

# A Computational Treatment of Superlatives

*Silke Scheible*



Doctor of Philosophy

Institute for Communicative and Collaborative Systems

School of Informatics

University of Edinburgh

2009

# Abstract

The use of gradable adjectives and adverbs represents an important means of expressing comparison in English. The grammatical forms of comparatives and superlatives are used to express explicit orderings between objects with respect to the degree to which they possess some gradable property. While comparatives are commonly used to compare two entities (e.g., “The blue whale is larger than an African elephant”), superlatives such as “The blue whale is the largest mammal” are used to express a comparison between a target entity (here, the *blue whale*) and its comparison set (the set of *mammals*), with the target ranked higher or lower on a scale of comparison than members of the comparison set. Superlatives thus highlight the uniqueness of the target with respect to its comparison set.

Although superlatives are frequently found in natural language, with the exception of recent work by (Bos and Nissim, 2006) and (Jindal and Liu, 2006b), they have not yet been investigated within a computational framework. And within the framework of theoretical linguistics, studies of superlatives have mainly focused on semantic properties that may only rarely occur in natural language (Szabolsci (1986), Heim (1999)).

My PhD research aims to pave the way for a comprehensive computational treatment of superlatives. The initial question I am addressing is that of automatically extracting useful information about the target entity, its comparison set and their relationship from superlative constructions. One of the central claims of the thesis is that no unified computational treatment of superlatives is possible because of their great semantic complexity and the variety of syntactic structures in which they occur. I propose a classification of superlative surface forms, and initially focus on so-called “ISA superlatives”, which make explicit the IS-A relation that holds between target and comparison set. They are suitable for a computational approach because both their target and comparison set are usually explicitly realised in the text.

I also aim to show that the findings of this thesis are of potential benefit for NLP applications such as Question Answering, Natural Language Generation, Ontology Learning, and Sentiment Analysis/Opinion Mining. In particular, I investigate the use of the “Superlative Relation Extractor“ implemented in this project in the area of Sentiment Analysis/Opinion Mining, and claim that a superlative analysis of the sort presented in this thesis, when applied to product evaluations and recommendations, can provide just the kind of information that Opinion Mining aims to identify.

## Acknowledgements

First of all, I would like to thank my supervisor Bonnie Webber for her guidance and support throughout my PhD years. Even after all the time I have spent working on this topic, I still can't find enough superlatives to describe just how exceptional her guidance has been. Her commitment and interest in my subject and her enthusiasm have been truly inspirational. She was always there, around the clock, with constructive feedback. I am truly honoured to have worked with her and full of gratitude for everything she has done for me.

Heartfelt thanks also go to my second supervisors Maria Milosavljevic and Jon Oberlander. I would further like to thank Johanna Moore and Caroline Heycock for their help and advice on my first year report, and Claire Grover and Mark Steedman for their helpful feedback after my DDD.

I would also like to thank my annotator Annette Leonhard for her time.

Finally, a big thank you also goes to Martin Durrell at the University of Manchester.

I gratefully acknowledge financial support in the form of an EPSRC Doctoral Training Award provided by the School of Informatics.

# Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text. This work has not been submitted for any other degree or professional qualification except as specified.

*(Silke Scheible)*

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	What are superlatives? . . . . .	1
1.2	Scope of the thesis . . . . .	2
1.2.1	Superlative relation extraction . . . . .	2
1.2.2	Challenges . . . . .	3
1.2.3	Elements of a computational treatment of superlatives . . . . .	6
1.3	Linguistic overview of superlatives . . . . .	6
1.3.1	Form of superlatives . . . . .	7
1.3.2	Non-comparative uses of superlatives . . . . .	8
1.3.3	Syntactic aspects of superlatives . . . . .	10
1.3.4	Semantic aspects of superlatives . . . . .	11
1.3.5	Subjective vs. objective superlatives and the role of text genre	12
<b>2</b>	<b>Background</b>	<b>15</b>
2.1	Applications . . . . .	15
2.1.1	Question Answering . . . . .	15
2.1.2	Ontology Learning . . . . .	21
2.1.3	Natural Language Generation . . . . .	23
2.1.4	Conclusion . . . . .	25
2.2	Previous work . . . . .	25
2.2.1	Jindal and Liu (2006) . . . . .	25
2.2.2	Bos and Nissim (2006) . . . . .	31
2.2.3	Conclusion . . . . .	35
<b>3</b>	<b>A New Classification of Superlatives</b>	<b>37</b>
3.1	Semantic classification of superlatives . . . . .	37
3.1.1	Type I: Property set comparisons (explicit/implicit) . . . . .	38

3.1.2	Type II: Relative set comparisons . . . . .	42
3.1.3	Type III: Subject-based set comparisons . . . . .	45
3.1.4	Type IV: Intensifiers . . . . .	48
3.1.5	Type V: Proportional quantifiers . . . . .	49
3.2	Classification of superlative surface forms . . . . .	50
3.2.1	Purpose . . . . .	50
3.2.2	Adjectival superlatives . . . . .	51
3.2.3	Adverbial class . . . . .	60
3.2.4	Idiomatic superlatives . . . . .	62
3.3	Summary . . . . .	65
<b>4</b>	<b>A Corpus of Superlatives</b>	<b>66</b>
4.1	Overview . . . . .	66
4.2	Annotation scheme . . . . .	67
4.2.1	Classification . . . . .	67
4.2.2	Target and Comparison Set Identification . . . . .	68
4.3	Pilot annotation study . . . . .	72
4.3.1	Data and procedure . . . . .	72
4.3.2	Results and discussion . . . . .	73
4.4	TextWiki corpus . . . . .	77
4.4.1	Motivation . . . . .	77
4.4.2	Structure of the corpus . . . . .	78
4.4.3	Distribution of superlative classes in TextWiki . . . . .	79
4.4.4	Data sets used in the experiments . . . . .	80
4.5	Conclusion . . . . .	80
<b>5</b>	<b>Identifying and Classifying Superlatives Automatically</b>	<b>82</b>
5.1	Overview . . . . .	82
5.2	Experiment 1: Identifying superlatives ( <i>SUP-Finder</i> ) . . . . .	82
5.2.1	Overview . . . . .	82
5.2.2	Method . . . . .	83
5.2.3	Results . . . . .	85
5.2.4	Discussion . . . . .	85
5.2.5	Conclusion . . . . .	87
5.3	Experiment 2: Classifying superlatives ( <i>SUP-Classifier</i> ) . . . . .	87
5.3.1	Overview . . . . .	87

5.3.2	Tools and approach . . . . .	88
5.3.3	Classification strategy: The <code>supClassifier</code> module . . . . .	91
5.3.4	Results . . . . .	111
5.3.5	Discussion . . . . .	113
5.3.6	Conclusion . . . . .	124
<b>6</b>	<b>Superlative Relation Extraction</b>	<b>127</b>
6.1	Overview . . . . .	127
6.2	Experiment 3: Identifying the target and comparison set of ISA-1 superlatives (T/CS-Identifier) . . . . .	127
6.2.1	Overview . . . . .	127
6.2.2	Method . . . . .	129
6.2.3	Identifying the comparison set span (CS-Identifier) . . . . .	130
6.2.4	Identifying the target span (T-Identifier) . . . . .	138
6.2.5	Results and discussion . . . . .	140
6.2.6	Conclusion . . . . .	148
6.3	Evaluation of the Superlative Relation Extractor (SRE) on Bos and Nissim (2006)'s data set . . . . .	148
6.3.1	Overview . . . . .	148
6.3.2	SUP-Classifier . . . . .	149
6.3.3	T/CS-Identifier . . . . .	150
6.3.4	Conclusion . . . . .	155
<b>7</b>	<b>Superlatives in Sentiment Analysis and Opinion Mining</b>	<b>157</b>
7.1	Overview . . . . .	157
7.2	Previous work and motivation . . . . .	158
7.3	Data . . . . .	160
7.4	Superlatives in Opinion Mining . . . . .	162
7.5	Experiment 4: Identifying ISA-1 superlatives in customer reviews . . . . .	165
7.5.1	Overview . . . . .	165
7.5.2	Results . . . . .	165
7.5.3	Discussion . . . . .	166
7.6	Experiment 5: Identifying T and CS of ISA-1 superlatives in customer reviews . . . . .	168
7.6.1	Overview . . . . .	168
7.6.2	Results . . . . .	169

7.6.3	Discussion . . . . .	169
7.7	Conclusion . . . . .	172
<b>8</b>	<b>Conclusion</b>	<b>174</b>
8.1	Summary and main findings . . . . .	174
8.2	Future research directions . . . . .	176
<b>A</b>	<b>Appendix to Chapter 2</b>	<b>179</b>
A.1	Superlatives in TREC . . . . .	179
A.1.1	TREC 2002 . . . . .	179
A.1.2	TREC 2003 . . . . .	180
A.2	“Other” question nuggets . . . . .	182
A.2.1	Superlative nuggets in TREC 2004/2005 . . . . .	182
A.2.2	Comparative nuggets in TREC 2004/2005 . . . . .	187
<b>B</b>	<b>Appendix to Chapter 3</b>	<b>188</b>
B.1	Definitions . . . . .	188
<b>C</b>	<b>Appendix to Chapter 5</b>	<b>190</b>
C.1	ISA1-Identifier: Analysis of false negatives . . . . .	190
C.2	ISA2-Identifier: Analysis of false negatives . . . . .	193
C.3	DEF-Identifier: Analysis of false negatives . . . . .	194
<b>D</b>	<b>Appendix to Chapter 6</b>	<b>197</b>
D.1	Description of PostMod rules . . . . .	197
D.2	T/CS-Identifier: Results . . . . .	204
<b>E</b>	<b>Appendix to Chapter 7</b>	<b>222</b>
E.1	Discussion of SUP-Classifier results on Hu and Liu (2004)’s corpus of customer reviews . . . . .	222
	<b>Bibliography</b>	<b>231</b>

# Chapter 1

## Introduction

### 1.1 What are superlatives?

The most common form of comparison in English is found in the use of gradable adjectives or adverbs. Many adjectives describe qualities that can be measured in degrees (such as size, beauty, age, etc). The term *gradability* refers to the possibility of placing the adjective on a scale of comparison, at a position higher or lower than the one indicated by the adjective alone. Similarly, gradable adverbs specify a degree to which the adverb applies, which may be either higher or lower than its absolute form does.

While gradability can be expressed by use of degree adverbs such as *barely*, *very*, or *highly*, adjectives and adverbs can also take different forms to indicate their position on a scale of comparison, as illustrated by these examples for adjectives:

- (1.1) a. English is *more interesting* than Chemistry.
- b. Maths is *less interesting* than Chemistry.
- (1.2) a. History is the *most interesting* subject at school.
- b. Physics is the *least interesting* subject.

The *comparative* form of an adjective or adverb is commonly used to compare two entities to one another with respect to a certain quality. For example, in (1.1) (a) and (b), two different subjects are ranked according to how interesting the speaker finds them. The *superlative* form of an adjective expresses the end spectrum of the scale, and is a powerful way to compare and contrast one object with a set of other objects. In

(1.2) (a) and (b), History and Physics are compared to all the other subjects at school. Milosavljevic (1999), who investigates the use of comparison in the description of entities, classifies superlatives as a type of *domain-based comparison*, whose discourse goal is to make explicit the relationship between certain entities within a domain. In particular, the purpose of such comparisons is to highlight the uniqueness or non-uniqueness of the entity in focus, and to prevent the hearer from forming misconceptions about the similarity of the class of entities. In the case of superlatives, the comparison takes place between an entity and its *contrast set*, which is the entity's complement in a member class, such as its parent class in a generalisation hierarchy (Milosavljevic, 1999). Superlatives can therefore be seen as a special type of domain-based comparison.

## 1.2 Scope of the thesis

This thesis proposes a computational treatment of superlatives, starting with the main challenges in automatically recognising and extracting their components. The present section provides an overview of the scope of this thesis.

### 1.2.1 Superlative relation extraction

From a computational perspective, superlatives are of interest because they express a comparison between a target entity (indicated by curly brackets) and its comparison set (square brackets), as in:

(1.3) {The blue whale} is the *largest* [mammal].

Here, the target *blue whale* is compared to the comparison set of *mammals*. My initial investigation of superlative forms showed that there are two types of relation that hold between a target and its comparison set:

- **Relation 1:** IS-A relation
- **Relation 2:** Superlative relation

The IS-A (or hypernymy) relation expresses the membership of the target in the comparison class (e.g. its parent class in a generalisation hierarchy). The superlative rela-

tion specifies a property which all members of the set share, but which the target has the highest (or lowest) degree or value of. Both of these relations are of great interest from a relation extraction point of view, and in **Chapter 2**, I discuss their use in applications such as Question Answering and Natural Language Generation.

That a computational treatment of superlatives is a worthwhile undertaking is also supported by the frequency of superlative forms in ordinary text: In a 250,000 word subcorpus of the WSJ corpus I found 602 instances (which amounts to roughly one superlative form in every 17 sentences), while in the corpus of animal encyclopaedia entries used by Milosavljevic (1999), there were 1059 superlative forms in 250,000 words (about one superlative form in every 11 sentences). These results show significant variation in the distribution of superlatives across different text genres (cf. also discussion in Section 1.3.5).

For an interpretation of comparisons, two things are generally of interest: What is being compared, and with respect to what this comparison is made. Given that superlatives express set comparisons, a computational treatment should therefore help to identify:

1. The target and comparison set
2. The type of superlative relation that holds between them

As the following section shows, this task is far from straightforward.

## 1.2.2 Challenges

### Classification of superlatives

One of the central claims of the thesis is that no unified computational treatment of superlatives is possible because of their great semantic complexity and the variety of syntactic structures in which they occur. Compare for example:

- (1.4) The blue whale is the *largest* mammal.
- (1.5) Of all pubs in Edinburgh, the White Hart Inn is the *oldest*.
- (1.6) Breast cancer is the *most frequently* diagnosed cancer among women.
- (1.7) This hotel is located in the *most privileged* area of the Costa del Sol.

(1.8) Orders will be shipped the *most quickly* when they are placed Monday or Tuesday.

(1.9) The Royal Mile is *busiest* during the Festival season.

All of the above examples contain a superlative form (highlighted in italics). However, they differ not only in their syntactic structure, but also in the way in which they express a comparison. The top examples contain a clear-cut comparison between a target item and its comparison set: *the blue whale* vs. *mammal* in (1.4); *the White Hart Inn* vs. *all pubs in Edinburgh* in (1.5), *breast cancer* vs. *cancer among women* in (1.6). Although examples (1.7)-(1.9) also involve comparisons, their targets and/or comparison sets are not that straightforward to identify. For example, sentence (1.7) does not specify the name of the target item (which would be the name of the most privileged area of the Costa del Sol), because it is not needed for the sentence to make sense. Other cases, like (1.8) and (1.9), do not involve a comparison between *entities*, but rather between different *states* or *parts* of one single entity. For example, in (1.9) we understand that of all times in the year, the Royal Mile is busiest during the Festival season. Thus, the comparison does not involve the Royal Mile and a set of other places, but how busy it is at different times in the year.

**Chapter 3** investigates these semantic types of superlative comparison in more detail, and proposes a classification of surface forms. For the superlative extraction task, I will initially focus on cases like (1.4), which I call “ISA superlatives” because they make explicit the IS-A relation that holds between target and comparison set (cf. Relation 1 in Section 1.2.1). They are a good initial focus for a computational approach because both their target and comparison set are explicitly realised in the text (usually, though not necessarily, in the same sentence). Common surface forms of ISA superlatives involve the verb “to be” (1.10)-(1.12), appositive position (1.13), and other copula verbs or expressions (1.14) and (1.15):

(1.10) The blue whale is the *largest* mammal.

(1.11) The blue whale is the *largest* of all mammals.

(1.12) Of all mammals, the blue whale is the *largest*.

(1.13) The *largest* mammal, the blue whale, weighs more than 30 elephants.

(1.14) The ostrich is considered the *largest* bird.

(1.15) Mexico claimed to be the *most peaceful* country in the Americas.

ISA superlatives are also the most frequent type of superlative comparison, with 176 instances in the WSJ subcorpus mentioned above (ca. 30% of all superlative forms), and 350 instances in the corpus of animal encyclopaedia entries used by Milosavljevic (1999) (ca. 33% of all superlative forms).

### Superlative relation

In addition to the challenges in identifying the target and comparison set of a superlative, the interpretation of the term “superlative relation” is also not straightforward. While the superlative form is generally taken to express a *dimension* on which the members of the set are compared (e.g. size, quality, or weight), it is not always possible to pinpoint this dimension: Superlatives are often ambiguous or vague in what dimension they express. For example, *biggest* in (1.16) expresses the property *size*, but it is unclear whether it refers to geographical space or population. An additional problem is represented by cases like (1.17). In contrast to (1.16), where *biggest* expresses a property which all members of the comparison set share (i.e. size), this is not the case in (1.17): The superlative does not entail that the target “its price” has the property “size”.

(1.16) The *biggest* city in Germany is Berlin.

(1.17) Its *biggest* advantage is its price.

It is clear that the “superlative relation” between *advantage* and *its price* in (1.17) is drastically different from the relation between *city in Germany* and *Berlin*. Detecting the dimension expressed by a superlative and interpreting the superlative relation between target and comparison set will require a detailed semantic analysis which is outside the scope of this work. Therefore, in the current study the task of identifying the “superlative relation” only involves the identification of the superlative keyword (such as *biggest* in 1.16) but not its dimension (e.g. *size*). The italicised forms in sentences (1.10)-(1.15) are further examples of what is meant by “superlative relation” in this work.

### 1.2.3 Elements of a computational treatment of superlatives

The proposed superlative extraction task can be seen as consisting of three subtasks:

- **TASK 1:** Decide whether a given sentence contains a superlative form
- **TASK 2:** Given a sentence containing a superlative form, identify what type of superlative it is (initially: ISA superlative or not?)
- **TASK 3:** For set comparisons, identify the target and the comparison set, as well as the superlative relation

In order to deal with these tasks automatically, a corpus is required where superlatives are annotated with their class membership, and where the target and comparison set strings of all instances classified as ISA are marked up. **Chapter 4** describes an annotation scheme for these two tasks, which has been tested and evaluated on 500 tokens of superlatives. In addition to providing a platform for investigating superlatives on a larger scale, Chapter 4 also introduces a new text-based Wikipedia corpus (“TextWiki”) which is especially suitable for linguistic research.

On the basis of the annotated Wikipedia corpus, **Chapters 5 and 6** describe three experiments: The first aims to identify all superlative-containing sentences in a given text (solving TASK 1). The second deals with the classification of superlatives according to the classes of surface forms proposed in Chapter 3 (TASK 2). For superlatives classified as ISA, Experiment 3 then tackles the identification of the target and comparison set spans (TASK 3). Finally, **Chapter 7** investigates the use of the resulting “Superlative Relation Extractor“ in the area of Sentiment Analysis/Opinion Mining, and claims that superlative analysis of the sort presented in this thesis, when applied to product evaluations and recommendations, can provide just the kind of information that Opinion Mining aims to find.

## 1.3 Linguistic overview of superlatives

Descriptions of superlatives as a linguistic construction can be found in all English grammar books, from simple grammars for learners of English to comprehensive grammars for linguists. This section provides a first overview of superlative constructions, which draws upon insights from three sophisticated reference grammars of British En-

English: A Comprehensive Grammar of the English Language by Quirk et al. (1985), the Cambridge Grammar of English by Huddleston and Pullum (2002), and the Longman Grammar of Spoken and Written English (Biber et al., 1999).

### 1.3.1 Form of superlatives

Generally, superlatives can be formed in two different ways: *inflectionally* or *analytically*. In the first case, the inflectional suffix *-est* is appended to the base form of the adjective or adverb, while in the second case they are preceded by the analytical markers *most/least*. Which of these options is used depends on the nature of the adjective/adverb. A common assumption is that the choice between inflection and periphrasis is related to the length of the base form. Quirk et al. (1985), for example, classify adjectives according to their number of syllables, stating that monosyllabic adjectives more typically form superlatives by inflection (e.g. *lowest*, *nicest*, *smartest*), while trisyllabic or longer adjectives can usually only be formed analytically (*most beautiful*, *most difficult*). Disyllabic adjectives are generally said to occur in both the analytical and inflectional forms (*politest/most polite*). For each group, Quirk et al. list a group of exceptions, and note that most adjectives that have inflectional superlatives can also be formed analytically. It has been argued that the reason why no clear paradigm can be stated for the formation of superlatives is that it may be affected by language change in progress (Bauer, 1994). Over the past centuries, English has gradually dropped most of its inflectional endings, thus developing towards being an analytical language. Some adjectives and adverbs, however, have irregular, lexicalised forms and are therefore unlikely to be affected by regularisation (Tables 1.1 and 1.2).

absolute	comparative	superlative
<i>good</i>	<i>better</i>	<i>best</i>
<i>bad</i>	<i>worse</i>	<i>worst</i>
<i>far</i>	<i>further</i>	<i>furthest</i>
	<i>farther</i>	<i>farthest</i>

Table 1.1: Irregular adjectival superlatives

While most adjectives and adverbs are gradable, there are also some that are *non-gradable*. This will be the case, inter alia, when the absolute form already indicates the highest or lowest position on a scale, for example *dead* or *awake*. Such adjectives

absolute	comparative	superlative
<i>well</i>	<i>better</i>	<i>best</i>
<i>badly</i>	<i>worse</i>	<i>worst</i>
<i>little</i>	<i>less</i>	<i>least</i>
<i>much</i>	<i>more</i>	<i>most</i>

Table 1.2: Irregular adverbial superlatives

do not usually have superlative forms, apart from in some poetic or intensifying uses (e.g. “it is deader than the *deadest* parrot”). The gradability of adjectives has been studied extensively in semantics, often in combination with investigations of their *polarity* or *semantic orientation* (e.g. Lyons (1977), Kennedy (1997), Klein (1980)). Polarity refers to the direction in which an adjective strays from its lexical field (Lehrer, 1974). Adjectives denoting desirable properties are usually conceived as positive, while those denoting undesirable properties are considered of negative polarity, for example *beautiful* vs. *ugly*.

### 1.3.2 Non-comparative uses of superlatives

Sometimes superlative forms do not have a superlative function. Huddleston and Pullum (2002) distinguish two uses of *most* as a non-superlative marker: intensifier (cf. also Quirk et al. (1985)), and proportional quantifier. *Most* as an *intensifier* has the role of a degree adverb with the meaning “very, extremely”. Consider the following examples taken from Huddleston and Pullum (2002):

- (1.18) Kim is a [most enthusiastic] supporter. [*intensifying*]  
 (1.19) This one is [most useful]. [*superlative or intensifying*]  
 (1.20) This one is [cheapest]. [*superlative only*]  
 (1.21) You are [most kind]. [*intensifying as salient reading*]

They argue that there is a clear semantic difference between (1.18) and the sentence “Kim is the most enthusiastic supporter I’ve come across”, as only the latter involves a comparison with a set of supporters. The two interpretations are distinguished by using the indefinite article *a* for the intensifying reading, and the definite article *the* for the superlative reading. In cases where there is no article, as in (1.19) and (1.21),

both the superlative and the intensifying reading are possible. In some cases one of the readings is more salient than the other, as for example in (1.21). The superlative of *kind* is commonly formed inflectionally (*kindest*), which might be the reason for why the marked form (formed analytically with *most*) gets the intensifying reading. Quirk et al. (1985) note that there is also a tendency to express intensification by using *most* with a preceding definite article, as for example in:

(1.22) Isn't she *the most beautiful woman*?

Here, *the most beautiful woman* means “an extremely beautiful woman”. They further draw attention to the fact that inflectional superlatives are also sometimes used in a similar fashion:

(1.23) Lucille wears *the oddest clothes*, my dear.

Huddleston and Pullum discuss a further non-superlative use of *most*, which is not dealt with by Quirk's grammar: Its use as a **proportional quantifier**. Consider the following examples from the Cambridge Grammar of English:

(1.24) Most people think he's guilty. [*proportional quantifier*]

(1.25) I agree with most of your points. [*proportional quantifier*]

(1.26) Kim had interviewed most candidates. [*superlative or proportional*]

(1.27) Kim had interviewed the most candidates. [*superlative only*]

In (1.24) and (1.25), *most* is not used as a superlative, but expresses a proportional quantification. In this context it means “more than half” or “the majority”. Huddleston and Pullum state that, again, ambiguous readings are possible, for example in (1.26). Here, the superlative reading compares the amount of people Kim had interviewed to the amount of people some other people had interviewed (making these other people the set against which Kim is compared). The proportional reading involves no comparison, but states that Kim had interviewed the majority of candidates. These readings can be distinguished by trying to insert the definite article: Only the superlative reading allows this (1.27).

### 1.3.3 Syntactic aspects of superlatives

From a structural perspective, one can distinguish between superlative forms that are bound to a noun phrase and those that are not. Huddleston and Pullum (2002) refer to these two categories as *incorporated superlatives* (Examples (1.28) - (1.33)) and *free superlatives* (Examples in (1.34) - (1.37)):

- (1.28) They rejected [the two best novels she has written].
- (1.29) Kim has [the most valuable collection of all].
- (1.30) This is [her most perfectly constructed novel].
- (1.31) Pat made [the most mistakes].
- (1.32) He offered me [the least valuable of the paintings].
- (1.33) [The most we can hope for] is a 2% rise.
- (1.34) She's [the candidate most likely to be elected].
- (1.35) These were the ones that the grown-ups laughed at loudest.
- (1.36) He's the least able to look after himself.
- (1.37) It was Jill who presented her case the most efficiently.

Huddleston and Pullum note that incorporated superlative forms either occur before the head of an NP (as in (1.28) - (1.31)) or are part of a so-called “*fused head*” construction, where the superlative and the head “merge” into one unit (1.32 and 1.33). The fused head theory is promoted by Huddleston and Pullum to account for the phenomenon where there is no obvious NP-head. Alternative explanations for this often found in syntax theory are NP-ellipsis and the null noun theory.<sup>1</sup>

A further discussion in this context deals with the presence or absence of the definite article. Quirk et al. (1985) describe superlative forms that are incorporated into an NP as “attributive”, and state that for these the definite article (or other definite determiner) is always required. Huddleston and Pullum put it slightly differently. They agree that definite determiners are always permitted and are obligatory most of the time, but admit that there are also cases where they are optional:

---

<sup>1</sup>For a detailed discussion see Huddleston and Pullum's chapter on noun phrases in the Cambridge Grammar of English (2002).

(1.38) The programme gives [(the) best results] if you begin before the age of thirty.

(1.39) The rebates should be given to those in [(the) greatest need].

(1.40) It was Kim who attracted [(the) most attention].

In free superlative phrases, on the other hand, *the* is fully optional. For example it could be inserted in (1.34) and (1.35) (although this is less common), and dropped from (1.36) and (1.37). However, they note that in the latter cases it is not possible to replace *the* with a possessive, as it does not function as determiner in the NP but is part of the superlative form.

### 1.3.4 Semantic aspects of superlatives

While there have been many studies on comparatives and gradable adjectives in theoretical linguistics (Kennedy (1997), Klein (1980)), relatively little has been done on superlatives. With the exception of an unpublished MA Thesis by Ross (1964), there has been no detailed study of the syntax of superlative phrases. However, one phenomenon on the level of semantics has been studied extensively over the last 20 years. Consider these two sentences described in the Cambridge Grammar of English (2002):

(1.41) Kim lives in the smallest house in England. [*absolute*]

(1.42) Of all members of the team, Kim had the most difficult job. [*relative*]

In (1.41), the comparison is with the set of houses in England, and doesn't involve anything outside the underlined NP. In contrast, the comparison in (1.42) involves Kim: What is being compared is how difficult Kim's job was to how difficult the jobs of the other members of the team were. These types of comparisons are commonly referred to as *absolute* and *relative* interpretation of superlative NPs, an observation first discussed by Ross (1964), and elaborated in a number of papers, most notably the ones by Heim (1985), Szabolsci (1986), and Heim (1999).

The distinction between absolute and relative interpretations is of great interest for semanticists because in some cases an ambiguity can arise that results in two readings of the same sentence. The often cited example for this is (Szabolsci, 1986):

(1.43) John climbed the highest mountain.

In the first reading, John climbed the highest of all mountains - in our world, this is Mount Everest. This is the absolute reading. However, one could also imagine a scenario where several people have climbed mountains, and of all the mountains people have climbed, John had climbed the highest one. Like (1.42), this interpretation is relative. What is of interest for semanticists is the question whether what we are dealing with here is a genuine structural ambiguity at the level of logical form (LF), or whether it is just a case of context-dependency. Most studies agree that the former is the case, and detailed explanations of this phenomenon have been proposed.

### 1.3.5 Subjective vs. objective superlatives and the role of text genre

The Longman Grammar of Spoken and Written English (Biber et al., 1999), which gives a corpus-based description of the grammar of English, shows that the distribution of inflected superlative forms (i.e. those with the suffix *-est*) varies quite drastically across registers. Of the four genres discussed (conversation, fiction, news, and academic prose), inflected superlatives occur most frequently in news (ca. 1400 instances per million words), and least frequently in spoken language (only 500 instances per million words). Fiction and academic prose occupy middle ground with about 700 and 800 instances per million, respectively.

Generally, two different types of superlatives can be distinguished: *Objective* (fact-based) and *subjective* (opinion-based) ones. The frequency with which these occur is likely to depend on the text genre. Consider the following examples:

(1.44) Fort William (Gaelic: An Gearasdan, “The Garrison”) is the *largest* town in the west highlands of Scotland.

(“*Fort William, Scotland*”, *Wikipedia*<sup>2</sup>)

(1.45) On Friday, analysts at Merrill Lynch reignited speculation that Bank of America, the world’s *biggest* bank, might be interested in acquiring Barclays, which is valued at 47.5 billion.

(*Citigroup hatches 950m Egg deal*, *The Sunday Times*<sup>3</sup>)

<sup>2</sup>[http://en.wikipedia.org/wiki/Fort\\_William,\\_Scotland](http://en.wikipedia.org/wiki/Fort_William,_Scotland)

<sup>3</sup>from *The Sunday Times* online (December 10, 2006) <http://business.timesonline.co.uk/article/0,,8209-2496141,00.html>

- (1.46) The blue bell is the *sweetest* flower / That waves in the summer air;  
Its blossoms have the *mightiest* power / To soothe my spirit's care.

(*"The Blue Bell"*, Emily Brontë)

- (1.47) N-E-S-T-L-E-S, Nestle's makes the very *best* chocolate.

(*Nestle advertisement*<sup>4</sup>)

- (1.48) I found this to be the *most poorly* written ill-researched piece of rubbish I have ever had the misfortune to lay my hands on.

(*Customer review of Dan Brown's novel The Da Vinci Code*<sup>5</sup>)

**Encyclopedias** are a good source for factual information, as the example in (1.44) illustrates. Encyclopedia entries should always express a neutral viewpoint.<sup>6</sup> **Newspaper articles** are another useful source for fact-based information (1.45), although they do allow for subjective views, depending on the individual paper, and on the section of the paper (e.g. editorials, personal columns, reviews and even sports news often allow for subjective views). Furthermore, subjective views are often expressed in quotations of other people. In **literary genres** like fiction or poetry, on the other hand, superlatives usually express fictional facts or opinions of characters. In addition, they are often used in an intensifying way, especially in poetry, for example to praise an object of affection (1.46).

Examples (1.47) and (1.48) represent two text genres which often express opinion. Superlatives are known to be a common instrument in **advertising** to exaggerate the function or effectiveness of products. In most cases they express vague or subjective opinions which cannot be proved true or false, as for example in (1.47). This method is usually referred to as *puffery* and constitutes a current legal issue in advertising: Courts have decided that puffery is legal because people should be prudent enough to be able to understand that the advert is subjective. Thus, advertisements may express almost any general subjective opinion. This, however, is a problem because there is only a thin line between puffery and deception, which is illegal (Preston 1996).

The opinions expressed in advertising tend to be highly unbalanced as they are mainly positive, which is not surprising because their purpose is of course to convince cus-

<sup>4</sup><http://www.nestle.co.uk>

<sup>5</sup><http://www.amazon.co.uk/Da-Vinci-Code-Dan-Brown/dp/0552149519>

<sup>6</sup>Wikipedia tries to achieve this by peer reviewing.

tomers to buy a product. Customer reviews as in (1.48), on the other hand, usually express more varied opinions. Volunteer reviewers tend to be very passionate about the product they are writing about, as it is either “so good” or “so bad” that they devote some of their time to describe it. In recent years customer reviews have become a popular genre in NLP due to the rising interest in areas such as Sentiment Analysis and Opinion Mining. Chapter 7 will discuss this in more detail.

# Chapter 2

## Background

This chapter aims to motivate the proposed computational treatment of superlatives. In the first part of the chapter, three possible applications of this PhD study are discussed, showing that the proposed work will be beneficial for a variety of NLP areas. The second part reviews two previous computational studies that attempt to deal with superlatives, and discusses their limitations in detail.

### 2.1 Applications

This section discusses possible applications of my PhD research within the areas of Question Answering, Ontology Learning, and Natural Language Generation. For each of these I will discuss in what ways they could benefit from an automatic treatment of superlatives. Furthermore, in a later chapter I will demonstrate an application of this work in the area of Sentiment Analysis/Opinion Mining (Chapter 7).

#### 2.1.1 Question Answering

##### Overview

Question Answering (QA) is a kind of information retrieval, where questions asked in natural language are handled by retrieving answers from a knowledge base. Generally, two types of QA are distinguished: Closed domain QA vs. open-domain QA. Closed-domain QA systems often deal with questions about a certain topic and use

structured or semi-structured data as their knowledge base (such as databases or ontologies). Open-domain QA systems, on the other hand, rely on unstructured data such as large natural language corpora or the World Wide Web, and attempt to deal a wide variety of questions and topics.

Over the last decade, research in QA has mainly focused on open-domain systems. Current systems are expected to deal with a wide variety of questions, the most important of which are factoid questions (e.g. *What is the capital of France?*), list questions (e.g. *What movies did Julianne Moore appear in?*), and definition questions (e.g. *What is pragmatics?*). The annual TREC Conference (**T**ext **R**etrieval **C**onference)<sup>1</sup> features a question answering track in which current systems' performance on different question types is assessed. The knowledge base used in this evaluation is the AQUAINT corpus,<sup>2</sup> which consists of around 375 million words of 1999 American English newswire text.

### Factoid questions

The present work could be beneficial to QA in two different ways, the first of which will be discussed in this section. Bos and Nissim (2006) found that superlatives are fairly common in naturally-occurring questions, with a frequency of about 1 per 25 questions (4%). This is reflected in their presence in the TREC QA question sets: In TREC 2002, 36 out of 500 questions contained superlatives (ca. 7%), and in 2003, they occurred with a frequency of 29 out of 500 (ca. 6%). TREC questions in those years mainly focused on factoid questions; in 2003, 413 of the 500 questions were factoids (ca. 83%). With the exception of one, all of the questions containing superlatives were of the factoid type.

The superlative questions of 2002 and 2003 are listed in Appendix A.1.1 and A.1.2, respectively. A striking proportion of them have the verb “to be” as their main verb: Of 65 questions altogether, this applies to 45 (69%). Examples are:

- (2.1) What is Africa's *largest* country?
- (2.2) What is the *deepest* lake in the world?
- (2.3) What roller coaster is the *fastest* in the world?

---

<sup>1</sup><http://trec.nist.gov/>

<sup>2</sup><http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2002T31>

(2.4) What is the *oldest* national park in the U.S.?

Of the other 20 questions, over one third do not involve regular superlatives but idiomatic expressions such as “Who won the Oscar for best actor in 1970?” This means that, effectively, 78% (45 out of 58) of the questions in TREC 2002/2003 describe a comparison involving the verb “to be”. This provides further motivation for an initial focus on superlatives where the target and comparison set stand in an “IS-A” relation, as discussed in 1.2.2.

If the proposed system could interpret a question like (2.1) by extracting a vector like (*target, superlative relation, comparison set, restrictions on comparison set*), i.e. in this case (“blank”, *largest, country, in Africa*), then the corpus could be parsed for sentences that contain superlative constructions with the same vector, and the “blank” target slot could be filled in. Although one could look for the answer to (2.1) (and (2.2) through (2.4)) by only searching for the string “is Africa’s largest country” and see what occurs in subject position in the resulting snippets, this method would not be able to account for the same variety of structures as the proposed method (consider e.g. “the largest country in Africa”, “the largest of all countries in Africa”). The proposed system therefore promises to be a useful help in finding answers for the majority of questions in TREC 2002/2003, and for a substantial proportion of questions asked in online question logs.

Linden (2008) demonstrated a first practical application in this area of the work described in this thesis. In her investigation of superlatives in question logs, she used the superlative classification introduced in this thesis and developed a program to find answers to superlative questions that belong to the ISA-1 class (“SuperAns”). The best performing version of SuperAns was able to extract a correct answer in the top two ranked answers for 17 out of 20 test questions.

### Definition (or “other”) questions

In 2004, the TREC QA track became topic-based, which meant that questions were no longer independent from each other but were clustered around a particular target, such as for example “the band Nirvana”:<sup>3</sup>

(2.5) <target id = “11” text = “the band Nirvana”>

<sup>3</sup>Adapted from [http://trec.nist.gov/data/qa/2004\\_qadata/QA2004\\_testset.xml](http://trec.nist.gov/data/qa/2004_qadata/QA2004_testset.xml)

<q id = “11.1” type=“FACTOID”> Who is the lead singer/musician in Nirvana?

<q id = “11.2” type=“LIST”> Who are the band members?

<q id = “11.3” type=“FACTOID”> When was the band formed?

<q id = “11.4” type=“FACTOID”> What is their *biggest* hit?

<q id = “11.5” type=“LIST”> What are their albums?

<q id = “11.6” type=“FACTOID”> What style of music do they play?

<q id = “11.7” type=“OTHER”> Other

Interestingly, the number of questions involving superlatives is drastically smaller in TREC 2004, with only 2 out of 351 questions in total (< 1%), even though the majority of questions is still factoid (230 altogether; ca. 66%). The reason for this decline is probably directly related to the introduction of topic-based questions: Questions containing superlatives often ask for a specific target, which in many cases happens to be a named entity (cf. (2.1)-(2.4) above). Thus, given a set of entities, the purpose of such questions is to identify the member of the comparison set with the highest or lowest degree of a certain property (denoted by the superlative form). As topic-based question sets already specify a target, factoid questions are restricted to asking about particular facts concerning this target, rather than for the target itself. Years 2005 and 2006 show only a slight increase of superlatives, with 12 out of 530 and 12 out of 567, respectively (ca. 2%). This means that the system proposed in this paper can only be of marginal help for answering factoid questions in current TREC QA competitions.

There is, however, a different way in which superlatives can be useful in topic-based scenarios. Rather than considering whether a *question* contains superlatives, one could also look at what sort of *answers* may involve superlative constructions. Compare:

(2.6) A: The Nile is the *longest* river in the world.

(2.7) Q: **What** is the *longest* river in the world?

(2.8) Q: **What** is the Nile?

Starting from an answer string like the one in (2.6), two different kinds of question can be constructed. Firstly, one can formulate a question that has the subject of the above sentence as its answer, *the Nile*. This results in a superlative question like (2.7) (this specific one appears in TREC 2003). In addition, one can also formulate a question to ask for the complement, “the longest river in the world”, as in (2.8), which results in a definition question. This question type has received a considerable amount of attention

in the last few years (e.g. Prager et al. (2001), Hildebrandt et al. (2004)).

Since TREC 2004, definition questions have come in the form of “other” questions. As shown in (2.5), the current TREC test sets consist of a series of questions that are related to a specific topic (the “target”). These questions are either factoid or list questions, apart from the last one, which is an “other” question. This last question type requires systems to return as many nuggets as possible that are relevant for an understanding of the topic. The idea that superlatives might be useful with respect to answering “other” questions is based on the observation that superlatives are often used for the purpose of defining an entity: They not only place the entity in a generalisation hierarchy, but also distinguish it from the other members of its class (cf. Chapter 1.1).

To corroborate the claim that superlatives are useful for answering “other” questions, I undertook a small study which involved the TREC 2004 and 2005 “other” question nuggets (Scheible, 2007). These are snippets of text that occur in the AQUAINT corpus, which are used by TREC to evaluate complex questions. The assumption is that snippets of text can either be relevant for answering a question or not. The “other” question nuggets are relevant snippets, and are further divided up into “vital” or “okay” ones to distinguish between important relevant nuggets and not so important relevant nuggets. Each of these nuggets has been assessed individually by 10 human judges (for more information on this procedure, see Lin and Demner-Fushman (2006)).

The question of interest is whether there is a correlation between nuggets containing superlatives and nuggets being judged as “vital”. I first looked at the distribution of superlatives among them. Of 139 “other” questions (64 in TREC 2004 and 75 in TREC 2005) and 1338 nuggets in total, 69 nuggets contain a superlative. This means that on average half of all questions have at least one superlative nugget. An investigation of these superlative nuggets proved to be very interesting: Out the 69, 20 are judged to be “vital” by all 10 assessors (almost one third). Another 12 nuggets are judged as “vital” by 9 assessors. Altogether, about 86% (59 out of 69) of the nuggets containing superlatives are judged as “vital” by at least 5 judges (see Appendix A.2.1 for a complete list of superlative nuggets).

A finer-grained analysis of the superlative nuggets reveals further interesting results. With respect to the nature of the target of the superlative comparison, the class can be divided up into three subclasses: The first one contains those superlative nuggets where the target of the superlative construction coincides with the target of the question (this

will be referred to as type S1). The second one includes nuggets where the superlative target is a part of or closely related to the target of the question, or is part of the comparison set (S2). The third subclass contains those superlatives whose target is not closely related (or unrelated) to the target of the question (S3). Table 2.1 shows an example of each class:

type	target	nugget	Qid
S1	AARP	<i>Largest</i> seniors organization	5.6.3
S2	Florence Nightingale	Nightingale Medal <i>highest</i> international nurses award	33.5.1
S3	Kurds	Irbil <i>largest</i> city controlled by Kurds	51.4.14

Table 2.1: Examples of superlative nuggets

The subclass S1 contains 46 nuggets, S2 contains 15 nuggets, and S3 contains 8. Table 2.2 shows that 87% of all judgements of the S1 instances are “vital”, while the vital judgements of the members of S2 and S3 are much lower, with 59% and 38%, respectively. Furthermore, the table in Appendix A.2.1 shows that all 20 cases judged as “vital” by all 10 assessors are of type S1. The remaining 26 members are judged as “vital” by at least 5 assessors, apart from one case (superlative no. 52, target “tsunami”, judged as “vital” by 4 assessors, cf. Appendix A.2.1). Thus, superlative nuggets judged as “vital” by less than half of the assessors almost exclusively belong to the S2 or S3 classes.

	# instances	# “vital”	# “okay”	% “vital”	% “okay”
S1	46	400	60	87%	13%
S2	15	89	60	59%	40%
S3	8	30	48	38%	60%

Table 2.2: Ratings of the classes S1, S2, and S3

To put these results into perspective, I also investigated the number of occurrences of comparative forms in the “other” question nuggets (such as *greater* or *more interesting*). Although comparatives occur far more frequently in newspaper text than superlatives (Biber et al., 1999), there are only 5 relevant nuggets that contain comparative forms. Moreover, the judges’ assessments of these nuggets show them to be not very important, with only one of the five being judged as “vital” by more than 5 judges (see Appendix A.2.2).

To draw reliable statistical conclusions one would also have to analyse occurrences of superlatives in “unrelevant” nuggets. However, the above results strongly suggest that the presence of superlatives, and in particular S1 membership, is a good indicator of the importance of nuggets. Experiments carried out in the framework of TREC 2006 (Kaisser et al., 2006), however, showed that a simple approach based on superlative forms alone (i.e. without taking class membership in S1, S2 or S3 into account) does not improve the results: A simple algorithm that reranked the “other” questions output of Kaisser’s QA system by moving snippets that contain superlatives higher up in the list did not score higher than the original output. An analysis of the superlative instances reveals that this is due to a large amount of S2 and S3 superlatives, i.e. superlatives which are not directly related or unrelated to the target of the question. A similar simple technique was used by Ahn et al. (2004) and by Razmara and Kosseim (2007). All just looked for the presence of a superlative and raised the score without further analysing the type of superlative or its role in the sentence. This calls for a more sophisticated approach, where class S1 superlatives can be identified, which is what this research proposes.

## 2.1.2 Ontology Learning

### Overview

Ontologies are formal specifications of sets of concepts and the relationships that hold between them, often within a particular domain. They have become an important means for automatically structuring information, and have proved their usefulness in areas such as knowledge discovery and data mining, and in various areas of speech and natural language processing. A relatively recent area of application is the Semantic Web,<sup>4</sup> where ontologies are expected to play a central role in its organisation and functioning.

The 2004 W3C Recommendation of Web Ontology Language (OWL) discusses the uses and requirements of ontologies for the Semantic Web.<sup>5</sup> Their major role lies in representing the semantics of documents, and making them usable by web applications, thus making these more “intelligent”. The W3C further state that ontologies represent a useful way of structuring and defining the meaning of metadata terms, which are

---

<sup>4</sup><http://www.w3.org/2001/sw/>

<sup>5</sup><http://www.w3.org/TR/webont-req/>

in the process of being collected and standardized for the emerging Semantic Web. According to the W3C OWL guidelines, ontologies should encode descriptions for the following kinds of concepts:

1. Classes (general things) in the many domains of interest
2. The relationships that can exist among things
3. The properties (or attributes) those things may have

As the hand-crafting of such ontologies is a time-consuming and costly task, recent research has focused on learning ontologies automatically or semi-automatically from unstructured text, e.g. *TextToOnto* (Maedche and Staab, 2000) and *Ontolearn* (Navigli et al., 2003). These systems use linguistic analysis and machine learning to identify useful concepts and the relations between them.

### Superlatives in ontology learning

The work proposed in this PhD could extend ontologies with a new type of relation that has not previously been considered: The superlative relation. This is particularly interesting for set comparisons between entities that involve the verb “to be”, as for example in:

(2.9) The Great Dane is the *tallest* dog.

(2.10) The Chihuahua is the *smallest* dog.

(2.11) The St. Bernard is the *heaviest* dog.

As discussed in 1.2.1, two different kinds of relation are involved here. The first relation, commonly referred to as “IS-A” relation, expresses the membership of an entity in a particular class (here: the particular dog breed and the general “dog” class). The second relation, the superlative relation, specifies a property that is shared by all members of the class (here: *height*, *weight*), and which the target has the highest (or lowest) degree or value of. By extracting a triple consisting of the target entity, the comparison set, and the two relations that hold between them, existing “IS-A” ontologies can be extended with a new relation that holds between the target entity and its superordinate class in the “IS-A” hierarchy, the superlative relation.<sup>6</sup> Thus, the information in

---

<sup>6</sup>Identifying the underlying property according to which entities are compared does not form part of this work (cf. Chapter 1.2.2).

(2.9)-(2.11) could extend an existing “IS-A” ontology as illustrated in Figure 2.1.

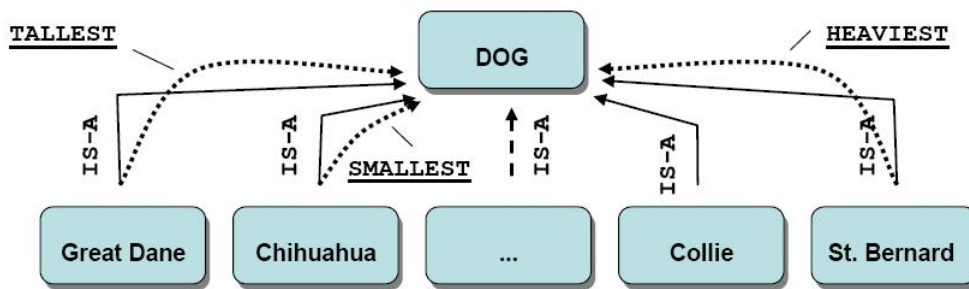


Figure 2.1: Superlative extension of an IS-A ontology

Once again, this provides motivation to initially focus the effort on superlatives involving the verb “to be”. Similar approaches, where existing ontologies are extended with new relations, are discussed in Schutz and Buitelaar (2005).

### 2.1.3 Natural Language Generation

#### Overview

The area of Natural Language Generation (NLG) is concerned with the task of automatically producing natural language text from a formal representation system such as a knowledge base or a logical form.

The purpose of many NLG systems is to provide a user with information about entities in a database. Milosavljevic (1999) investigates the use of comparison in the automatic generation of descriptions of entities, and states that comparison is an important tool for maximising the understanding of such descriptions. By comparing one entity to entities that users are already familiar with, the description can both build on the users’ existing knowledge of these entities, and prevent them from forming misconceptions. As discussed in Chapter 1.1, Milosavljevic classifies superlatives as a type of set complement comparison, where the comparison takes place between an entity and its contrast set, noting that this is a useful way of highlighting the uniqueness of the entity in focus.

## Superlatives in NLG

From an implementational perspective, not much has yet been done on superlatives in NLG. The reason for this is possibly that set complement comparisons can also be achieved by use of comparatives: Most systems that deal with comparison focus on a so-called “compare-and-contrast” schema, in which entities are compared pairwise (e.g. McKeown (1985), Milosavljevic (1999)). If this procedure is used to compare the target item to all other members of its contrast set, then this can be seen as having the same function as a superlative set comparison. Milosavljevic (1999) draws attention to the fact that this method is often used when there is no known name for the contrast set, or when the number of items in the comparison set is very small. Knott and Mellish (1997) describe an approach to generating set comparisons in the ILEX system, which uses defeasible rules, and allows for comparisons between a target entity and *most* members of a set.

None of these approaches explicitly use superlative forms for the purpose of set comparisons. Having a method of producing superlative comparisons would however clearly be useful, as the “compare-and-contrast” technique can end up in very lengthy descriptions for large comparison sets, which may confuse rather than help the reader. The structural and semantic analysis of superlatives undertaken in this study will provide useful information about the way in which superlative constructions could efficiently be used in generation.

One particular application of superlatives in natural language generation is in the generation of referring expressions (GRE), which is one of the most widely explored areas in NLG. It deals with the problem of establishing the content of an expression that uniquely identifies an intended referent. One of the main challenges in this task is that the generated expressions should sound as natural as possible. This means that they have to conform to a variety of constraints to avoid unwanted conversational implicatures (Grice (1975), Reiter (1990)). Up until recently, work in this area has mainly focused on generating referring expressions for identifying objects that are in the immediate focus of attention (e.g. Reiter and Dale (1992)). As such sets of objects are commonly rather small in number, the use of superlatives is not necessary, as other modifiers (e.g. adjectives like *big* or *blue*) can be used to achieve identification. In addition, objects in immediate focus can often be identified by description of their location.

In other scenarios, where the entity to be identified is not directly in focus, and where there is large set of potential confusors, the use of superlatives can be a useful method of identifying this item. A first step in this direction was made by van Deemter (2006), who examines the role of gradable properties in referring expressions from an NLG point of view. Although his work is on “immediate focus” entities as well, he finds that in human speakers’ use of referring expressions, superlative forms are often preferred over comparative ones, even when only comparing two things. The role of superlatives therefore seems to be more important than previously assumed. The proposed work can therefore prove very useful for determining the structure of natural sounding superlative referring expressions.

#### **2.1.4 Conclusion**

This section discussed potential applications of the proposed superlative extraction system in the areas of Question Answering, Ontology Learning, and Natural Language Processing. The application of superlatives in QA and Ontology Learning in particular showed the usefulness of superlatives that involve an “IS-A” relation between target and comparison set, thus motivating an intital focus on this type of superlative comparison. In a later chapter, an application in a further area in NLP will be demonstrated: Sentiment Analysis/Opinion Mining (Chapter 7).

## **2.2 Previous work**

### **2.2.1 Jindal and Liu (2006)**

#### **Goal**

In 2006, two papers were published by Jindal and Liu<sup>7</sup> that propose the study of *comparative sentence mining*. By “comparative sentence” Jindal and Liu denote sentences that express “an ordering relation between two sets of entities with respect to some common features” (2006b). Their interest lies mainly in subjective comparisons, which is why they focus on evaluative texts on the web, for example reviews, forum postings,

---

<sup>7</sup>I would like to thank Nitin Jindal and Bing Liu for making their data set available.

and blogs (cf. also Section 1.3.5). Comparative sentence mining is defined as a two-step task, where the first step is to identify comparative sentences in a given text, and the second one is to extract comparative relations from the identified sentences. In particular, they define the term “comparative relation” as follows:

*A comparative relation* captures the essence of a comparative sentence and is represented with the following:

(relationWord, features, entityS1, entityS2, type)

where *relationWord*: The keyword used to express a comparative relation in a sentence.

*features*: a set of features being compared.

*entityS1* and *entityS2*: Sets of entities being compared. Entities in *entityS1* appear to the left of the relation word and entities in *entityS2* appear to the right of the relation word.

*type*: *non-equal gradable*, *equative* or *superlative*.

(Jindal and Liu 2006b)

An *entity* in this context is defined as the name of a person, a product brand, a company, a location, or similar, under comparison in a comparative sentence, and a *feature* is taken to be a part or property of the entity that is being compared. Entities and features can only be nouns (including plural and proper nouns) or pronouns. Thus, for a sentence like “*Canon’s optics is better than those of Sony and Nikon*”, the system is expected to extract the following relation:

(better, {optics}, {Canon}, {Sony, Nikon}, *non-equal gradable*)

Jindal and Liu define the four different types of comparison *non-equal gradable*, *equative*, *superlative*, and *non-gradable*, as follows:

1. *Non-equal Gradable*: Relations of the type *greater* or *less than* that express a total ordering of some entities with regard to certain features. This type also includes user preferences.
2. *Equative*: Relations of the type *equal* to that state two entities as equal with respect to some features.
3. *Superlative*: Relations of the type *greater* or *less than* all others that rank one entity over all others.
4. *Non-Gradable*: Sentences which compare features of two or more entities, but do not explicitly grade them.

(Jindal and Liu 2006b)

Jindal and Liu only focus on the first three types, which they call “gradable comparatives”, as they express an explicit ordering between entities (2006b).

## Approach

Jindal and Liu’s treatment of the first task, namely identifying comparative sentences, is based on an approach that involves class sequential rules (CSR) and naïve Bayesian classification, and is described in Jindal and Liu (2006a). Extracting the comparative relations described above consists of two steps: First, all comparative sentences are classified as one of the three types *non-equal gradable*, *equative*, or *superlative* by using an SVM learner. The second step deals with the extraction of features, entities, and relation keywords for each of these types. For this task, Jindal and Liu propose a type of rules called *label sequential rules* (LSR). Label sequential rules are of the form “ $X \rightarrow Y$ ”, where  $Y$  is a sequence of items (e.g. words), and  $X$  is a sequence produced from  $Y$  by replacing some of its items with wildcards. As the goal of the comparative relation extraction task is to identify the items *entityS1*, *entityS2*, and *feature* in a given comparative sentence, LSR rules are generated by creating a sequence database based on the context of each of these items (using a radius of 4 words). To do this, initially all nouns and pronouns in comparative sentences are manually annotated with labels *entityS1*, *entityS2*, *feature*, or *non-entity feature*.<sup>8</sup> For every instance of the labels *entityS1*, *entityS2*, or *feature*, a separate sequence is added to the database. This sequence takes into account not only lexical and POS-based information of the context words of the respective label (called the ‘pivot’ of the sequence), but also includes information about the distance of the context words to their pivot label (also using a radius of 4, e.g. “r3” indicates a distance of 3 to the right of the pivot). In addition, there are indicators for the start and end of sentences. For example, in the comparative sentence “Canon/NNP has/VBZ better/JJR optics/NNS than/IN Nikon/NNP”, “Canon” is labelled as *entityS1* (\$ES1), “optics” as *feature* (\$FT), and “Nikon” as *entityS2* (\$ES2), and three separate sequences are put in the database, one using the label \$ES1 as pivot, and the other two using \$FT and \$ES2 as pivots, respectively. For instance, when considering \$ES1 (“Canon”) as pivot, the following sequence is added:

<sup>8</sup>This is based on the assumption that there is only one relation type in each sentence, and that entities and features are expressed via nouns or pronouns.

$$(\{\#start\} \{l1\} \{\$ES1, NNP\} \{r1\} \{has, VBZ\} \{r2\} \{better, JJR\} \{r3\} \{\$FT, NNS\} \{r4\} \{thanIN\})^9$$

This sequence considers four words to the right of \$ES1 (where {r2} indicates that the distance of the pair {better, JJR} to \$ES1 is 2), but as \$ES1 is the first word in the sentence, there is only a “start” indicator to its left, and no further context to the left is considered.

The sequence database is then used to generate LSR rules. First, frequent sequences are identified by using a formula to establish their “minimum support” (i.e. the fraction of instances in the input data set that match the pattern). For each of these, four LSR rules (of different confidence levels) are generated by allowing their respective pivot element (which is a noun or pronoun) to occur with any of the POS tags NN, NNP, NNS and PRP, thus introducing the required “wildcard” to the LSR rule. The rules are then applied to the comparative sentences in the data set to extract the components of the comparative relation.

## Data and results

The data consist of customer reviews, forum discussions, and random news articles taken from the Internet, and was labelled by two annotators (disagreement was resolved through discussion). Of 3248 sentences in total, 285 were labelled as *non-equal*, 110 as *equative*, and 169 as *superlative*. In total, the annotators labelled 488 instances of *entityS1*, 300 instances of *entityS2*, and 348 *features*. The overall F-score for the extraction task is 72%, a big improvement to the 58% achieved by their baseline system (which was the CRF system developed by Sarawagi 2004). According to the graph presented in Jindal and Liu (2006b), LSR achieved around 80% F-score for *entityS1*, about 70% for *entityS2*, and around 60% for *features*. It would have been interesting to see how the system performed on superlatives alone. However, no information is available on the F-scores of the individual types of comparison.

## Limitations

With superlatives making up almost one third of the comparative sentences, the results suggest that Jindal and Liu’s system represents a powerful way of dealing with

---

<sup>9</sup>sic.

superlatives computationally. However, a closer inspection of their approach, and in particular their data, reveals a number of problems.

The first problem concerns the definition of the *superlatives* class. It not only contains grammatical superlatives, but also other constructions that have superlative function, i.e. “relations of the type *greater* or *less than all others* that rank one entity over *all others*” (Jindal and Liu, 2006b). Examples of keywords used for this category are “top”, “only”, “one of the”, and “one of the few”. This definition is problematic because these keywords also frequently occur with a non-superlative interpretation. In addition, it is far from clear what is actually meant by “*all others*”. This may lead to difficulties in deciding which of the two classes a sentence belongs to. Consider for example (2.12), which uses a comparative form, indicating a “non-equal gradable” type, but could also be considered as belonging to the superlatives class according to Jindal and Liu’s definition:

(2.12) Canon’s optics is better than those of all other cameras of this price.

Another problem concerns the definition of the entity sets *entityS1* and *entityS2*. Membership in the two sets is not determined by semantics, but by relative position in the sentence: *entityS1* appears to the left of the relation word, and *entityS2* appears to the right of it (see definition above). This definition was probably derived from the consideration that the relation that holds between the two sets is asymmetric for both the *non-equal gradable* and the *superlative* classes. This means that, like in a mathematical “greater than” relation, the greater argument should occur on the left hand side of the relation symbol, and the smaller argument on the right hand side. However, unlike in maths, natural language does not always conform to such a strict ordering of constituents, and an asymmetric relation can manifest itself in a variety of surface forms. This particularly applies to the superlative relation. Consider for example the following sentences, which are semantically equivalent with respect to the *best* relation:

(2.13) The White Hart Inn offers the *best* wine selection of all pubs in Edinburgh.

(2.14) Of all pubs in Edinburgh, the *best* wine selection is offered at the White Hart Inn.

According to the above definition of *entityS1* and *entityS2*, such variation would result in different relation vectors for these sentences:

(2.13) (best, wine selection, The White Hart Inn, all pubs in Edinburgh)

(2.14) (best, wine selection, all pubs in Edinburgh, The White Hart Inn)

Obviously, such an approach runs into severe semantic problems when interpreting the output of the system: An entity's occurrence in the *entityS1* slot does not necessarily mean that, from a formal perspective, it occurs on the left hand side of the relation. Jindal and Liu's result of 100% precision for *entityS1* should therefore be considered with caution.

The problem just described could be easily solved by using a semantic definition of the entity sets, where *entityS1* is associated with the entity in focus (the *target* of comparison) and where *entityS2* represents the *comparison set*, i.e. the set of entities the target is compared to. However, a closer look at how instances of the superlative class are annotated in the gold standard data set reveals further problems. According to Jindal and Liu, for superlatives the *entityS2* slot is "normally empty" (2006b). Assuming that the members of *entityS2* represent the comparison set, this would mean that the comparison set is "normally empty", which is somewhat counter-intuitive. A look at the data and its annotation shows why this is the case:

(2.15) i think , apex is the best dvd player you can get for the price .

**entityS1:** apex, **entityS2:** –, **feature:** dvd player, **relation word:** best

(2.16) simply without a doubt the finest looking apex dvd player that i 've seen .

**entityS1:** –, **entityS2:** –, **feature:** looking apex dvd player, **relation word:** finest

(2.17) overall , the g3 delivers what must be considered the best image quality of any current > 4 megapixel digicams , from a detail , tonal balance and color response point of view .

**entityS1:** g3, **entityS2:** –, **feature:** image quality, **relation word:** best

In constructions where the superlative form is incorporated into an NP, Jindal and Liu consistently interpret the string following the superlative form as a *feature*. While this is appropriate for cases like (2.13) or (2.14), it does not apply to superlative sentences involving the copula verb "to be". Here, the NP denotes the comparison set rather than a feature (the feature is expressed by the superlative form itself, e.g. *best* represents the feature "quality" in 2.15). Moreover, in their gold standard annotation, the *entityS2* slot remains empty even in cases where the comparison set is explicitly mentioned (as in (2.17), where the comparison set consists of *current >4 megapixel digicams*).

A further major problem is that a number of important semantic aspects are completely disregarded. For example, important restrictions on the comparison set (or *feature*, in their terms), are not considered:

- (2.18) After Fiat , Coca-Cola is now the second biggest investor in Poland .  
**entityS1:** Coca-Cola, **entityS2:** –, **feature:** investor in Poland, **relation word:** biggest

This results in the inaccurate information that *Coca-Cola* (and not *Fiat*) is the biggest investor in Poland. Furthermore, negation is not taken into account:

- (2.19) the software is somewhat nice to look at , using it is not the easiest .  
**entityS1:** software, **entityS2:** –, **feature:** using it, **relation word:** easiest

- (2.20) the scroll button is n't the best , as it sometimes can be hard to select .  
**entityS1:** scroll button, **entityS2:** –, **feature:** –, **relation word:** best

Although the task of comparative sentence mining described by Jindal and Liu (2006b) proposes to extract both the target and comparison set of superlatives and the relations that hold between them, the data reveals some crucial problems in the approach. There are inconsistencies in the annotation of the material, and important semantic aspects of superlatives are not taken into account, which makes the reliability of the output produced by the system problematic.

## 2.2.2 Bos and Nissim (2006)

### Goal

In contrast to Jindal and Liu (2006b), Bos and Nissim's approach to superlatives is explicitly semantic (Bos and Nissim, 2006). They draw attention to the variety of syntactic and morphological forms that superlatives can take, and distinguish between four classes, which they define with the following examples:<sup>10</sup>

1. **attributive:** *the strongest dividend growth*
2. **predicative:** *its rates will be among the highest*
3. **adverbial:** *free to do the task most quickly*

<sup>10</sup>I would like to thank Malvina Nissim and Johan Bos for making their data set available.

#### 4. **idiom:** *who won the TONY for best featured actor?*

(Bos and Nissim, 2006)

Bos and Nissim describe an implementation of a system that can automatically detect superlatives, and determine the correct comparison set for attributive cases, where the superlative form is incorporated into an NP.

#### **Approach**

Bos and Nissim's system, called DLA (Deep Linguistic Analysis), uses a wide-coverage parser (C&C, cf. Clark and Curran (2004b)) to produce semantic representations (DRS, cf. Bos et al. (2004)) of superlative sentences, which are then exploited to identify attributive cases and to select the comparison set among these. Bos and Nissim note that the output of the parser is not always sufficient to construct a meaningful semantic representation. In particular, it is unable to handle NP post-modification of the superlative on the one hand, and possessive NPs preceding the superlative construction on the other, resulting in wrong semantic analyses for both of these issues. The first problem is caused by an attachment problem, which Bos and Nissim illustrate with the example "*the largest toxicology lab in New England*", where the parser attaches the modifier to the NP node rather than N. The PP *in New England* is assigned the CCG category NP\NP instead of N\N, which implies that the comparison set consists of *toxicology labs*, rather than *toxicology labs in New England*. In the second problem, which Bos and Nissim illustrate with the example "*Jaguar's largest shareholder*", the parser outputs a derivation where "largest" is first combined with "shareholder", and then with the possessive construction (rather than the other way round), which also produces a wrong semantic interpretation. This problem is analysed in detail in a recent paper by Bos (2009).

Bos and Nissim address these issues by using post-processing rules to alter the CCG derivation output by the parser, resulting in four different versions of their DLA (Deep Linguistic Analysis) system. The first one uses the unmodified DRS output of their system (DLA 1). The second and third versions build on DLA 1 by adding post-processing rules to the CCG derivations, to deal with NP post-modification of the superlative on the one hand (DLA 2), and with possessives preceding the superlative on the other (DLA 3). The fourth version, which is shown to perform best, combines both post-processing rules into one system (DLA 4).

For the comparison set determination, Bos and Nissim developed two baseline systems. The first one takes the first word following the superlative as the beginning of the comparison set, and the first word tagged as NN.\* in that sequence as the end. Their second baseline takes the first word after the superlative as the beginning of the comparison set, and the end of the sentence as the end (excluding the final punctuation mark). They note that this approach is likely to generate comparison sets much wider than required.

### **Data and results**

As a first step, a training and test corpus was compiled from parts of the Wall Street Journal, the Glasgow Herald, and sets of questions from the TREC QA exercise and from natural language query logs. These were annotated with gold standard information about the type of superlative (one of the four classes above) and the span of the comparison set. The annotation was carried out by two trained linguists, with high inter-annotator agreement: 0.963 F-score on what sentences contain superlatives, 0.974 F-score on what type of superlatives these were, and 95.31% agreement on the span of the comparison set for attributive cases. The attributive class is by far the most common: Of 3,045 superlatives in the corpus (which amounts to around one superlative in every 25 sentences or questions), 89.1% are attributive, 6.9% are predicative, 3.0% are adverbial, and 0.9% are idiomatic.

The task of superlative detection achieves between 0.84-0.93 F-score. With the exception of DLA 1, all systems for comparison set determination (DLA 2, 3, and 4) outperform the baseline systems on text documents. DLA 4 has the best results with an accuracy of 69%-83% (depending on the subcorpus). However, on questions it competes with the baseline: While it outperforms the baseline on TREC questions, it performs worse on questions in the Excite corpus. This can be explained by the fact that the parser's model for questions was trained on TREC data. Excite questions, on the other hand, are difficult to parse as they are frequently ungrammatical (Bos and Nissim, 2006).

## Limitations

The results are clearly very promising and show that comparison sets can be identified with high accuracy. However, their work only represents a first step towards the goal of the present work. Firstly, Bos and Nissim only focus on the comparison set, but disregard the question of what the comparison set is actually compared to. Consider for example:

(2.21) *wsj00 1690*

(...) *Scope: 3-7*

The *oldest* bell-ringing group in the country , the Ancient Society of College Youths , founded in 1637 , remains male-only , a fact that 's particularly galling to women because the group is the sole source of ringers for Britain 's most prestigious churches , St. Paul 's Cathedral and Westminster Abbey

(Bos and Nissim 2006)

Apart from the superlative keyword *oldest*, the only information this example provides is that the comparison set spans from word 3 to word 7 (underlined in the example). However, what would be interesting to know is that the comparison set acts as a contrast set in a comparison with a target item, and that this target item spans from word 9 to word 14 (*the Ancient Society of College Youths*).

Furthermore, Bos and Nissim (2006) identify the *span* of the comparison set without further analysis of the constituents of the resulting string. This means that no information is available about what sort of entities the comparison set actually contains. By treating the set in the example above as a single string, the following useful information is lost:

1. The class the target belongs to, here expressed by an IS-A relation: [the Ancient Society of College Youths] IS-A [bell-ringing group]
2. The way in which this class is restricted, for example here in location: We do not consider *all* bell-ringing groups, but only those [in the country]

Finally, the focus on *attributive* cases is only motivated by structure, but not by their semantics. Consider the examples described in Chapter 1.2.2:

(2.22) The blue whale is the *largest* mammal.

(2.23) This hotel is located in the *most privileged* area of the Costa del Sol.

(2.24) Of all pubs in Edinburgh, the White Hart Inn is the *oldest*.

According to Bos and Nissim’s classification, only (2.22) and (2.23) would be classified *attributive*, and they would be treated by the system in exactly the same way. However, only the former describes an explicit comparison between a target item and a comparison set. Example (2.23), on the other hand, does not involve an explicit comparison at all. Furthermore, examples like (2.24) would be excluded from consideration. It would be classified as *predicative*, even though it is clearly also a set comparison with an explicit target and comparison set.

### 2.2.3 Conclusion

The limitations of the two studies discussed above show that for a useful computational treatment of superlatives as outlined in Chapter 1.2, a thorough linguistic investigation of their syntax and semantics is required. In the chapters to come, a study of superlatives is carried out which analyses the semantics and surface forms of superlatives in detail. This analysis will serve as a basis for a number of experiments which aim to overcome the limitations of the approaches by Jindal and Liu (2006b) and Bos and Nissim (2006).

The first experiment automatically identifies superlatives in free text using a pattern-based approach, and is shown to outperform a POS-based one. The second experiment attempts to classify superlatives according to their surface forms, with the aim of identifying cases that are particularly suitable for an extraction task with both target and comparison set explicitly mentioned in the sentence (“ISA-1” class). Focusing on such cases represents an improvement over the approaches discussed above: Jindal and Liu do not distinguish between different superlative classes at all, which causes major problems in the interpretation of their results, especially with regard to the *entityS2* slot (representing the comparison set), which according to Jindal and Liu is “normally empty” (cf. Section 2.2.1). Bos and Nissim, on the other hand, focus on attributive cases, but disregard the presence of a target of comparison. They furthermore exclude

“fused head” cases from consideration (cf. example (2.24)). The third experiment described in this thesis aims to overcome these limitations by addressing the task of “Superlative Relation Extraction” in terms of identifying both target and comparison set spans of ISA-1 superlatives.

Experiments 2 and 3 employ a rule-based approach based mainly on tag sequences and dependency relations (using the output of the C&C tools, cf. Clark and Curran (2004a)). In contrast to Jindal and Liu (2006b), emphasis is placed on achieving high precision, which means that if results are obtained they can be considered reliable. A rule-based approach was chosen over a machine learning one for a number of reasons. First of all, the small size of the development and test data sets represents a problem for a learner, in particular because of the low frequency of some superlative types. A second consideration was that the tools used to obtain the tags and dependency relations will have been optimised to correctly tag frequently occurring phenomena in its target text type, in order to achieve the highest possible performance score. As superlatives are relatively low frequency phenomena, with most types occurring far down the end of low frequency patterns (part of “the long tail”), even a relatively high-performance tagger like C&C may perform poorly at tagging them, because it will make little difference to the tagger’s overall performance score. The current approach allows for highly flexible and fine-tuned rules which can take these factors into account wherever necessary, and in Chapter 6.3 I am able to demonstrate that this approach outperforms Bos and Nissim’s system on comparison set spans on their own data. Whether a learner is able to beat the results achieved by the current rule-based system remains open for future research.

## Chapter 3

# A New Classification of Superlatives

Superlatives differ not only in their semantics, i.e. in the ways in which they express comparisons, but they also occur in a wide variety of syntactic structures. For a computational treatment of superlatives that has as its goal to extract the main components of a superlative comparison (i.e. information about the target and comparison set), it is of vital importance to distinguish between the different types of comparison that superlatives can express, and to identify the ones whose components lend themselves most readily for extraction.

The latter task is complicated by the fact that the relation between semantic type of comparison and syntactic surface form is far from straightforward, especially since surface forms are often ambiguous between different semantic readings. The present chapter therefore aims to: 1.) Provide a general overview of different types of comparisons that superlatives can express; and 2.) Propose an exhaustive classification of superlative surface forms that provides a useful platform both for a computational extraction task, and for a closer theoretical investigation of the semantics of superlative comparisons.

### 3.1 Semantic classification of superlatives

This section investigates the main ways in which superlatives are used to express comparisons. Of particular interest is what role the superlative plays in the comparison, and how difficult it is to determine the target and comparison set, which is crucial for an extraction task. The overview does not intend to provide an exhaustive list of semantic

types, and only adjectival superlatives are considered. While some of the proposed types of comparison can be adapted to describe adverbial superlatives as well, their semantics are more complex and will not be discussed in detail here. The five semantic types described in this section are:

- Type I: Property set comparisons (3.1.1)
- Type II: Relative set comparisons (3.1.2)
- Type III: Subject-based set comparisons (3.1.3)
- Type IV: Intensifiers (3.1.4)
- Type V: Proportional quantifiers (3.1.5)

### 3.1.1 Type I: Property set comparisons (explicit/implicit)

A superlative NP like “the youngest rescue dog” describes a property set comparison between a target T and a comparison set CS, as illustrated in Figure 3.1.

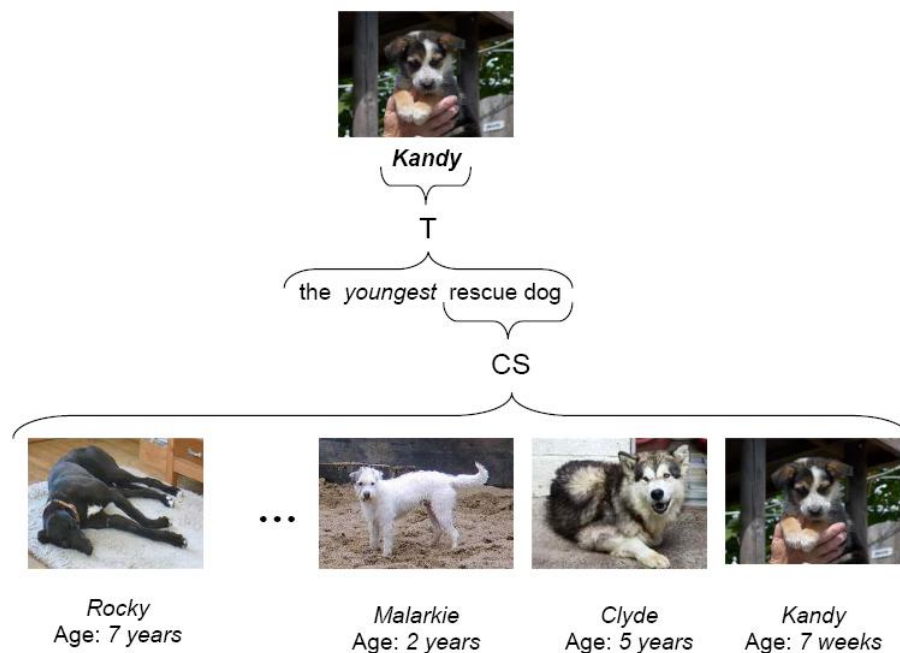


Figure 3.1: Target and comparison set of a property set comparison

The members of CS are being compared to one another with respect to a certain property  $p$ , in this case *age*. (The property is usually specified implicitly in terms of the

base adjective of the superlative form, here: “*young*”). Property set comparisons single out T as the member of the comparison set with the highest or lowest value of  $p$  (in the example, the dog “*Kandy*” takes the lowest value of the property *age*). Thus, two kinds of relations hold between T and CS, of which the former is inferred and the latter is asserted:

1. **IS-A relation**  $\implies$  IS-A(*Kandy*, *rescue dog*)
2. **Superlative relation**  $\implies$  youngest(*Kandy*, *rescue dog*)

Figure 3.2 illustrates this type of comparison further. In a sentence like “The whale shark is the largest fish”, the target “the whale shark” is compared to its contrast set, i.e. its complement in the CS: *basking sharks*, *stingrays*, *clownfish*, *goldfish*, etc. All of the members of the CS “fish” have a property “size” (indicated by the “HAS property” relation), which in turn has a value, here measured in *kg* (*up to 13,600 kg* in the case of the *whale shark*). It is these values that are compared in a property set comparison, and the superlative form expresses that the target has the highest or lowest value compared to the members of the contrast set. Note that in contrast to the example in Figure 3.1, where the members of the comparison set have real-world referents (specific dogs), the comparison set “fish” in Figure 3.2 contains *classes* of fish rather than individual fish. In cases where classes are compared, the property values often represent maximum values or are averaged over the whole class.<sup>1</sup>

Property set comparisons generally appear in one of two surface forms: explicit or implicit. In order to understand the difference between them, a distinction needs to be made between an entity (or entity set) *outside* the text that is being described and/or specified, the superlative description that may be asserted (or believed or speculated etc.) to hold of an entity, and finally text that provides a separate unique designator for an entity (or entities). I will use the following terminology:

- **T (Target):** The target entity outside the text that is being described
- **CS (Comparison Set):** The set of entities outside the text that are being compared
- **TD (Target Designator):** Text that provides a separate unique designator for T
- **CSD (Comparison Set Designator):** Text that provides a separate unique des-

---

<sup>1</sup>In many cases it is impossible to determine the exact values that are compared, which particularly applies to superlatives expressing *subjective* properties.

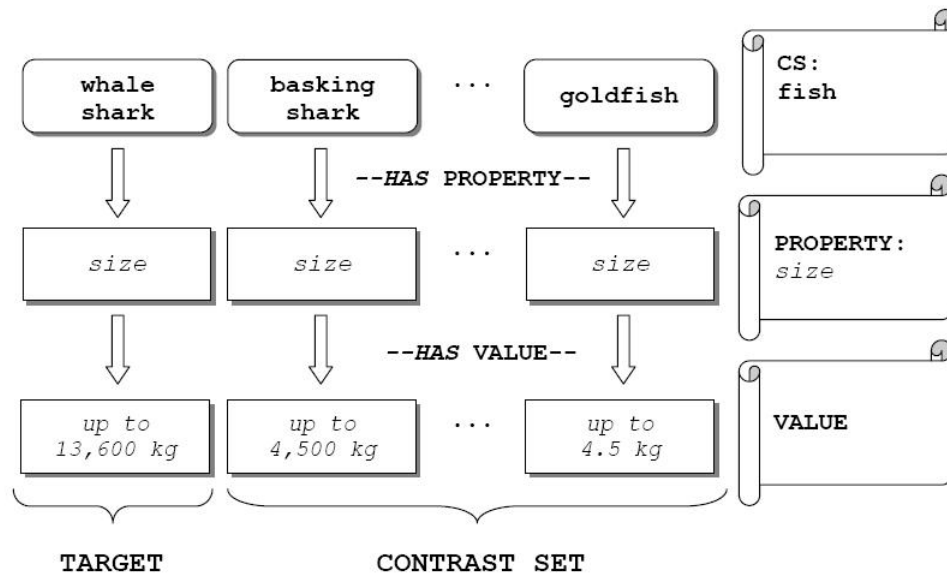


Figure 3.2: Property set comparison: “The whale shark is the *largest* fish.”

ignator for CS

- **S (Superlative Description):** The superlative description that is asserted to hold of T

Thus, in the example “Kandy is the youngest rescue dog”, “Kandy” is the Target Designator TD, “the youngest rescue dog” is S, and there is some entity T in the world that TD designates and S is asserted to hold of. The set of entities (outside the text) which is being compared (and which T is part of) is CS, which in the example is designated by the CSD “rescue dog” (i.e. a substring of S). To illustrate this further, in an example like

(3.1) I bought a new book. It is the best book I’ve ever read.

a constructed designator like “the new book I bought and just told you about” is TD, “the best book I’ve ever read” is S, and there is again some entity T in the world that TD designates and S is asserted to hold of. The comparison set CS is designated by a substring of S, “book I’ve ever read” (CSD).

In the **explicit surface form** of property set comparisons, both the superlative description S and the target designator TD are realised in the text, as for example in the sentence “The whale shark is the largest fish”. This is done by using a surface con-

struction that indicates equivalence between TD and S, for example by using the verb “to be” (or another copula verb), resulting in an SVC structure (3.2), or by appositive position of either TD or S (3.3).<sup>2</sup>

(3.2) {The whale shark} is [the *largest* fish].

(3.3) [The *largest* fish], {the whale shark}, weighs up to 13.6 tonnes.

An alternative way of expressing the IS-A relation between TD and S explicitly is by use of a so-called “fused head” construction (Huddleston and Pullum, 2002), where the superlative and the NP head of S “merge” into one unit. The NP head (which denotes the comparison set) can be indicated by use of an “of” PP phrase (3.4 and 3.5), or is implied elsewhere in the context (3.6):

(3.4) The whale shark is the *largest* of all fish.

(3.5) Of all fish, the whale shark is the *largest*.

(3.6) (There are many fish.) The whale shark is the *largest*.

In the **implicit surface form** of property set comparisons, only S is given: There is no separate designator TD for target T. In these cases, the superlative description S is used as a referring expression to denote the target. This happens when there is no particular “name” for the target, or when a designator of the target is not important for the given context. Consider for example:

(3.7) On [the *longest* day of the year], people all over Finland go to their summer cottages to celebrate Midsummer.

(3.8) Lisa bought [the *most expensive* watch in the shop].

Although the longest day of the year traditionally has a name in Finland (*Juhannus*), and although the watch that Lisa bought is a *Rolex*, it is not necessary to mention these designators explicitly to get the meaning of the sentences across. Both examples could be turned into *explicit* surface forms as follows:

(3.9) On [the *longest* day of the year], {Juhannus}, people all over Finland go to their summer cottages to celebrate Midsummer.

(3.10) Lisa bought [the *most expensive* watch in the shop], {this Rolex}.

<sup>2</sup>In the examples, TD is surrounded by curly brackets, while S appears in square brackets.

### 3.1.2 Type II: Relative set comparisons

The semantics of **relative set comparisons**, as in example (3.11), are complex as they build on property set comparisons, thus involving two interdependent set comparisons. Consider:

(3.11) Of all the  $[fishermen]_{CSD1}$ ,  $\{Peter\}_{TD1}$  caught the *largest*  $[fish]_{CSD2}$ .

The relative set comparison takes place between *Peter* (designator TD1) and the comparison set designated by *fishermen* (CSD1). The comparison is called *relative*, because its comparison set does not directly involve the superlative form and therefore only takes place relative to a second comparison involving the superlative form: the property set comparison with the superlative description *the largest fish* (S) and the comparison set designator “fish” (CSD2).

Again, there are implicit and explicit forms of the property set comparison. With *implicit* forms, there is no designator TD2 for the target T2 of the superlative comparison (as in 3.11). If it was a *bass* that Peter caught, then the property set comparison could be turned into an explicit surface form by adding “this bass” as an apposition to the end of the sentence:

(3.12) Of all the  $[fishermen]_{CSD1}$ ,  $\{Peter\}_{TD1}$  caught the *largest*  $[fish]_{CSD2}$ ,  $\{\text{this bass}\}_{TD2}$ .

Figure 3.3 aims to illustrate how the relative set comparison introduces another layer to the property set comparison. Important here is the relation between the two comparison sets, *fishermen* (CS1) vs. *fish* (CS2), which is expressed by the main verb of the clause (*caught*). Each member of CS1 stands in a “catch” relation with (at least) one instance of CS2.<sup>3</sup> Crucially, the comparison set CS2 is restricted by CS1, which means that every restriction imposed on CS1 also affects CS2: We do not consider all fish that exist on this planet (or even in the lake where the fishermen went fishing), but only the ones that are *caught* by one of the fishermen. This means that the whale shark, the “largest fish” in Figure 3.2, is excluded from the comparison set.

<sup>3</sup>There is not necessarily a one-to-one relation between the two sets. Some fishermen may have caught several fish: For example, (3.11) could be followed by “*He also caught the smallest one*”, without contradiction. The *verb* relation between CS1 and CS2 could therefore be said to be surjective (or *onto*), as every element in CS2 is hit by some argument from CS1, but not necessarily injective (*one-to-one*).

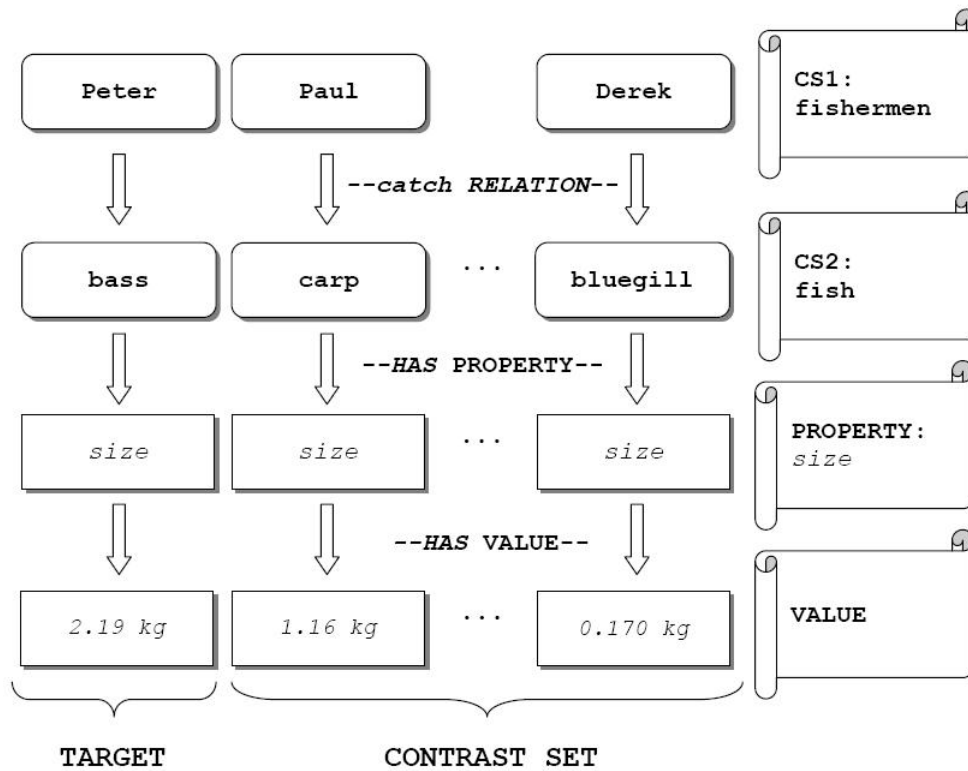


Figure 3.3: Relative set comparison: “Of all the fishermen, Peter caught the *largest* fish.”

Relative set comparisons are of interest for superlative relation extraction as they are often used in “feature comparisons”, where a set of entities is compared with respect to the (adjectival) property of a particular feature they all share. Consider for example the following excerpts from a camera review titled “*Canon SX100IS, Sony DSC-H3 or Panasonic Lumix DMC-TZ3?*”:

- (3.13) {The TZ3}<sub>TD1</sub> has the *fewest* [pixels]<sub>CSD2</sub> (7.1MP vs. 8.1MP and 8.3MP), which isn’t a big deal. Oddly enough, {the TZ3}<sub>TD1</sub> also has the *largest* [CCD]<sub>CSD2</sub> (1/2 .35” vs. 1/2.5”) which, coupled with fewer MPs will help low-light performance.
- (3.14) {The SX100}<sub>TD1</sub> has the *best* [macroability]<sub>CSD2</sub>, focusing in as close as 1cm.
- (3.15) {The H3}<sub>TD1</sub> has the *strongest* [flash]<sub>CSD2</sub> of [the 3]<sub>CSD1</sub>, followed by the TZ3 and then the SX100.

In Chapter 7, I describe how the analysis of superlatives types described in this chapter can be used for just this sort of information extraction.

Relative set comparisons often occur in an SVO structure, where the subject is realised by TD1, the verb expresses the relation that holds between the members of CSD1 and CSD2, and O is occupied by the superlative description S.<sup>4</sup> While TD1 and CSD2 are crucial constituents of relative set comparisons, TD2 is usually not made explicit (even though it could be, for example by apposition, as mentioned above). More importantly, CSD1, which from an extraction point of view is an essential component of the comparison, is not necessarily made explicit in the immediate context. Unlike the comparison set of a property set comparison, which is commonly incorporated in the superlative phrase (as in 3.2 and 3.3), CSD1 can only occur freely in the context, either explicitly by use of an “of” PP phrase like in (3.11) or (3.15), or implicitly like in (3.13) and (3.14), where CSD1 has to be inferred from the context by the reader.

This ellipsis of CSD1 makes relative set comparisons problematic candidates for superlative relation extraction, not just because relations remain incomplete as a result, but also because omission of CSD1 leads to the (in theoretical linguistics) much-discussed ambiguity between absolute and relative readings of superlatives (cf. Chapter 1.3.4). Compare:

---

<sup>4</sup>However, other structures are also possible, for example: “Of all residents in Saint Tropez, Pam owns the villa with the *largest* balcony”, where S is part of a PP phrase.

- (3.16) Lisa bought the *most expensive* [watch in the shop], {this Rolex}. [*property set comparison*]
- (3.17) Of [the people who bought watches at the shop today], {Lisa} bought the *most expensive* [watch]. [*relative set comparison*]
- (3.18) Lisa bought the *most expensive* watch. [*ambiguous*]

Example (3.18) is ambiguous between a relative set comparison and an implicit property set comparison. A resolution of this type of ambiguity has to deal with very complex semantic issues (cf. Szabolsci (1986), Heim (1999)).

### 3.1.3 Type III: Subject-based set comparisons

This type of superlative comparison focuses on a fixed subject that is not itself the target of the superlative comparison. Like relative set comparisons, subject-based set comparisons are semantically complex, and it is often difficult to determine what constitutes target and comparison set. Consider:

- (3.19) In July, Mercury is (the) *brightest* [on the first day of the month].
- (3.20) Left/right confusion is *most dangerous* [in car parks].
- (3.21) The human foot is *narrowest* [at the heel].

Here, we are not dealing with a comparison between a target entity and its comparison set as such, but rather with a comparison of a property intrinsic to a subject at different “states”, usually along the dimensions of time or space, and commonly indicated by PPs or adverbial phrases/clauses. For example, in (3.19), the subject is *Mercury*, the intrinsic property is *brightness*, and the states are different points in *time*, namely different days in July. There is no explicit designator of the CS, but could be paraphrased as “Mercury on different days in July”, while the target could be paraphrased as “Mercury on the first day of July”, as illustrated in Figure 3.4. In (3.20), the subject is *left/right confusion*, the property is expressed by *dangerous*, and the states are different *locations* in which the subject can be found (here: *in the car park* vs. other places you can go by car). In addition, rather than comparing the state of the subject in different locations, subject-based set comparisons can also compare different locations *within* or *parts of* the subject itself, as for example in (3.21), where the comparison set includes different parts of the subject “human foot”.

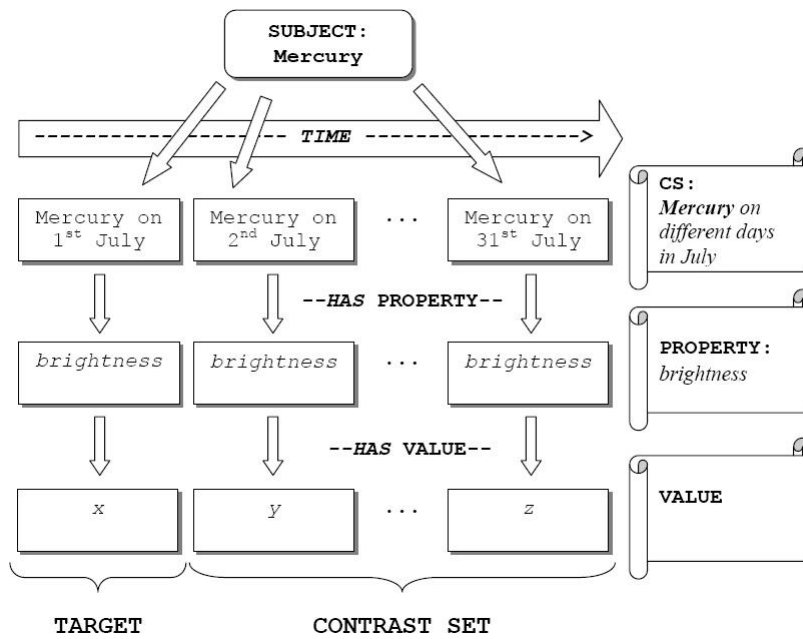


Figure 3.4: Subject-based set comparison I: “In July, Mercury is *brightest* on the first day of the month.”

Subject-based set comparisons can also involve NP-bound superlatives, as for example in:

(3.22) John sells the *largest* number of cars [on Saturdays].

Here, the comparison is between different week days on which John sells cars.

A second group of subject-based set comparisons involves a special kind of superlative relation between a subject and a comparison set, usually expressed by superlatives derived from fixed *adjective + preposition* combinations such as *famous for*, *afraid of*, or *popular with*, as in:

(3.23) Shakespeare is *most famous for* [his plays].

(3.24) People are *most afraid of* [terrorism].

(3.25) Camping is *most popular with* [young people].

In contrast to property set comparisons, the base adjective of the superlative form does not express a *property* that all members of the comparison set share, but a *relation* between the subject and a set of entities. For example, in (3.23) the subject is *Shakespeare*, the comparison set consists of everything Shakespeare is famous for (which

apart from *his plays* may also include *his sonnets*, *his prose*, *his writing style*, *his use of metaphors*, etc.), and the subject stands in a “famous for” relation with each member of the set. The comparison is illustrated in Figure 3.5: Instead of comparing a value associated with a property that all members of the CS share, here the *intensity of the relation* between subject and each CS member is compared, and the superlative form identifies the relation with the strongest intensity. In (3.23), the pair [*Shakespeare*, *his plays*] has the strongest “famous for” relation.

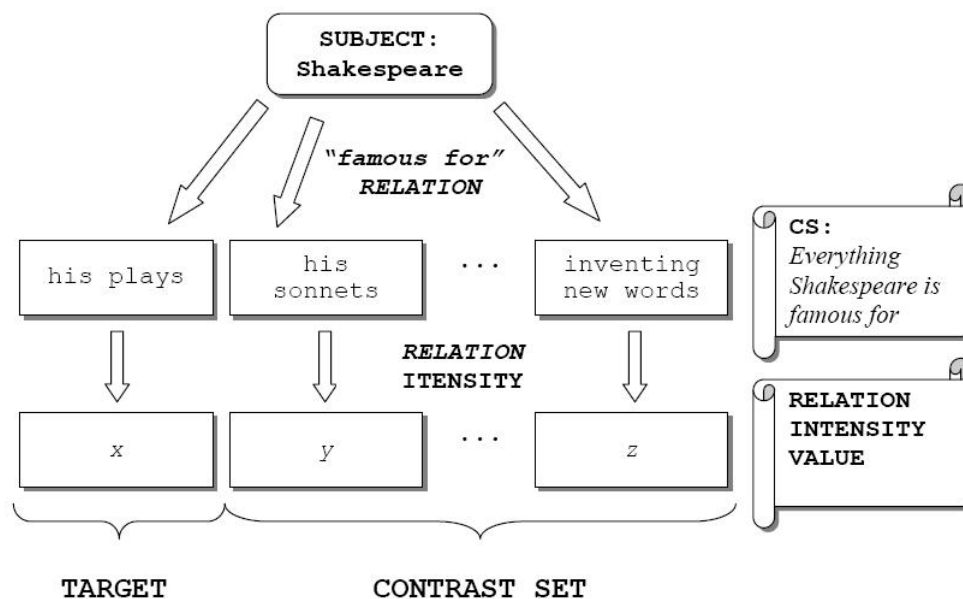


Figure 3.5: Subject-based set comparison II: “Shakespeare is *most famous* for his plays.”

Again, an explicit designator of the CS is not obligatory in the sentence, and like in the case of relative set comparisons, omission of a CSD string can lead to ambiguity:

(3.26) Mercury is *brightest* on the first day of July.

(3.27) UK consumers buy the *largest* numbers of video games within Europe.

(3.28) People in big cities are *most afraid* of terrorism.

Example (3.26) has at least three different readings. The first one is essentially a property set comparison which compares the brightness of *Mercury* on the first day of July to the brightness of other planets on the first day of July. The second one is a subject-based set comparison where the brightness of *Mercury* is compared on the first day of different months (*July* vs. *January*, *February*, etc.), and the third one is also a subject-

based set comparison where the brightness of *Mercury* is compared on different days in July (the first vs. the second, the third, etc.) In (3.27), one reading involves a relative set comparison between *UK customers* and other customer groups, while another one compares the numbers of video games that UK consumers buy *within Europe* to the numbers that they buy elsewhere in the world (subject-based set comparison).

In written language, different readings can only be disambiguated by the context, while in spoken language, intonation plays an additional role, with focus (expressed by means of pitch accent) being placed on the target of comparison. For example, in a sentence like (3.28), focus could be placed on the following words, introducing different readings (the focused word is highlighted in bold):

- People in big cities are *most afraid* of **terrorism**. (As opposed to other things they are afraid of)
- People in **big** cities are *most afraid* of terrorism. (As opposed to medium-sized or small cities)
- **People** in big cities are *most afraid* of terrorism. (As opposed to animals or other beings)

Due to their great level of ambiguity, subject-based set comparisons would represent a major challenge for an automatic superlative relation extraction.

### 3.1.4 Type IV: Intensifiers

As discussed in Chapter 1.3.2, superlatives can take on the role of a degree adverb with the meaning “very, extremely” to intensify the meaning of the base adjective. One can distinguish between two different cases, one where the superlative clearly does not involve a set comparison, and one where a set comparison still takes place. Compare:

(3.29) The chateau is very beautiful and has a *most interesting* history.

(3.30) This *most beautiful* chateau has a very interesting history.

(3.31) The Ritz is the *most amazing* hotel and the service was incredible.

(3.32) I am *most tired* of Amazon recommending products to me based on products I have purchased as gifts.

Superlatives with indefinite articles (like in 3.29) and demonstrative determiners (like in 3.30) are used as intensifiers and do not involve a set comparison (Huddleston and Pullum, 2002). While superlatives preceded by a definite article can also be used in an intensifying way (3.31), they still involve a set comparison, which is however subjective (see Chapter 6 for further discussion). For superlatives with no determiner usually both superlative and intensifying readings are possible, but one of the readings may be more salient than the other. In (3.32) an intensifying reading (“I am very tired of Amazon...”) competes with a subject-based set comparison reading, where the comparison set contains things the speaker is tired of. In written language, the ambiguity may be resolved by the surrounding context, while in spoken language, intonation and pitch accents may again play an additional role: Placing focus on the superlative seems to evoke the intensifying reading, while focus on *Amazon* evokes a superlative reading that compares *Amazon* to other companies that recommend products in the same way.<sup>5</sup>

### 3.1.5 Type V: Proportional quantifiers

As discussed in Chapter 1.3.2, the word *most* can express a proportional quantification with the meaning “more than half” or “the majority”:

(3.33) *Most* books have more than 5 pages.

(3.34) I have already eaten *most* of the biscuits I bought yesterday.

In the above cases *most* does not have a superlative meaning and involves no comparison. However, ambiguity can arise between proportional quantification and superlative readings, as in:

(3.35) I’ve read *most* books written by Virginia Woolf. [The only ones I haven’t read are *Jacob’s Room* and *The Waves*.]

(3.36) I’ve read *most* books written by Virginia Woolf. [I have read 7, Sally has read 5, Peter has read 2, and Toby none.]

While *most* in (3.35) is understood as a proportional quantifier, meaning “the majority of”, it has a superlative reading in (3.36) (a relative set comparison).

<sup>5</sup>This hypothesis would however have to be validated by spoken language experiments with native speakers of English.

## 3.2 Classification of superlative surface forms

### 3.2.1 Purpose

The previous section has given an overview of the semantic complexity of superlative comparisons and has shown that superlative surface forms are often ambiguous between different semantic readings. As there is no one-to-one relationship between semantic types and surface forms, a classification of the latter should attempt to separate superlatives up into useful classes that allow a detailed study of the relationship between the respective surface forms and their semantic readings. As this thesis focuses on a computational treatment of superlatives, another purpose of the classification is to group together superlative surface forms that are suitable for superlative relation extraction (as described in Chapter 1.2.1), which involves:

1. Identifying the target and comparison set strings
2. Determining the type of superlative relation that holds between target and comparison set

Of the semantic types discussed in Section 3.1, it may be easiest to satisfy the above requirements for property set comparisons, as the superlative form is usually bound in an NP whose head designates the CS, thereby making the CS an obligatory component of the sentence. Furthermore, focusing on property set comparisons with an *explicit* surface form ensures that both the comparison set string (CSD) and the target string (TD) are extractable. Although these strings only act as *designators* for the T and CS entities outside the text, and may therefore require knowledge of the context (especially where they involve pronouns or are incomplete), from now on the target designator (TD) will be referred to as target T, and the comparison set designator (CSD) as the comparison set CS.

Concerning point 2.), the superlative relation in property set comparisons has the role of identifying the target as the member of the comparison set with the highest or lowest value of a property  $p$  (which is expressed via the superlative form). While relative set comparisons and subject-based set comparisons also express interesting superlative relations, their extraction is complicated by the fact that there is often no designator of the CS in the sentence, and by the great amount of ambiguity that arises when the CS is omitted. Relative set comparisons are however common in product reviews and

hence of potentially great interest for Opinion Mining.

The proposed classification distinguishes between three main groups, containing eight different classes altogether:

- Adjectival superlatives: ISA, DEF, INDEF, FREE (3.2.2)
- Adverbial superlatives: ADV (3.2.3)
- Idiomatic superlatives: IDIOM, PP, PROP (3.2.4)

Figure 3.6 shows a binary tree which shows how class membership is determined for a particular superlative form. Each node represents a particular question, and each leaf stands for one of the eight classes. Depending how each successive question is answered, the superlative is labelled with the leaf that is reached. The adjectival group contains the four right-most leaves, which describe set comparisons as discussed in 3.1. Adverbial superlatives form a class of their own (ADV). The third group contains superlative instances that are of an idiomatic nature, represented by the IDIOM, PP and PROP leaves in the tree. The following sections provide an overview of each of the classes.

### 3.2.2 Adjectival superlatives

As Figure 3.6 illustrates, the first syntactic feature that is used to distinguish between different adjectival superlatives is whether they are bound in an NP (in which case they are referred to as “superlative NP comparisons”) or occur freely in a sentence (Question Q5). If they occur freely in the sentence, they are classified as FREE. Among the superlative NP comparisons, a further syntactic feature (represented by Question Q6) distinguishes between definite and indefinite (classified as INDEF) superlative NPs. Among the definite cases, Question Q7 further distinguishes between explicit (classified as ISA) and implicit surface forms (classified as DEF).

#### 3.2.2.1 ISA (Explicit superlative NP comparisons)

##### Surface forms

The ISA class contains definite superlative NP comparisons where both target and comparison set are explicitly mentioned in the context by use of an “ISA” surface construction (see below). ISA membership is determined by the following criteria (for

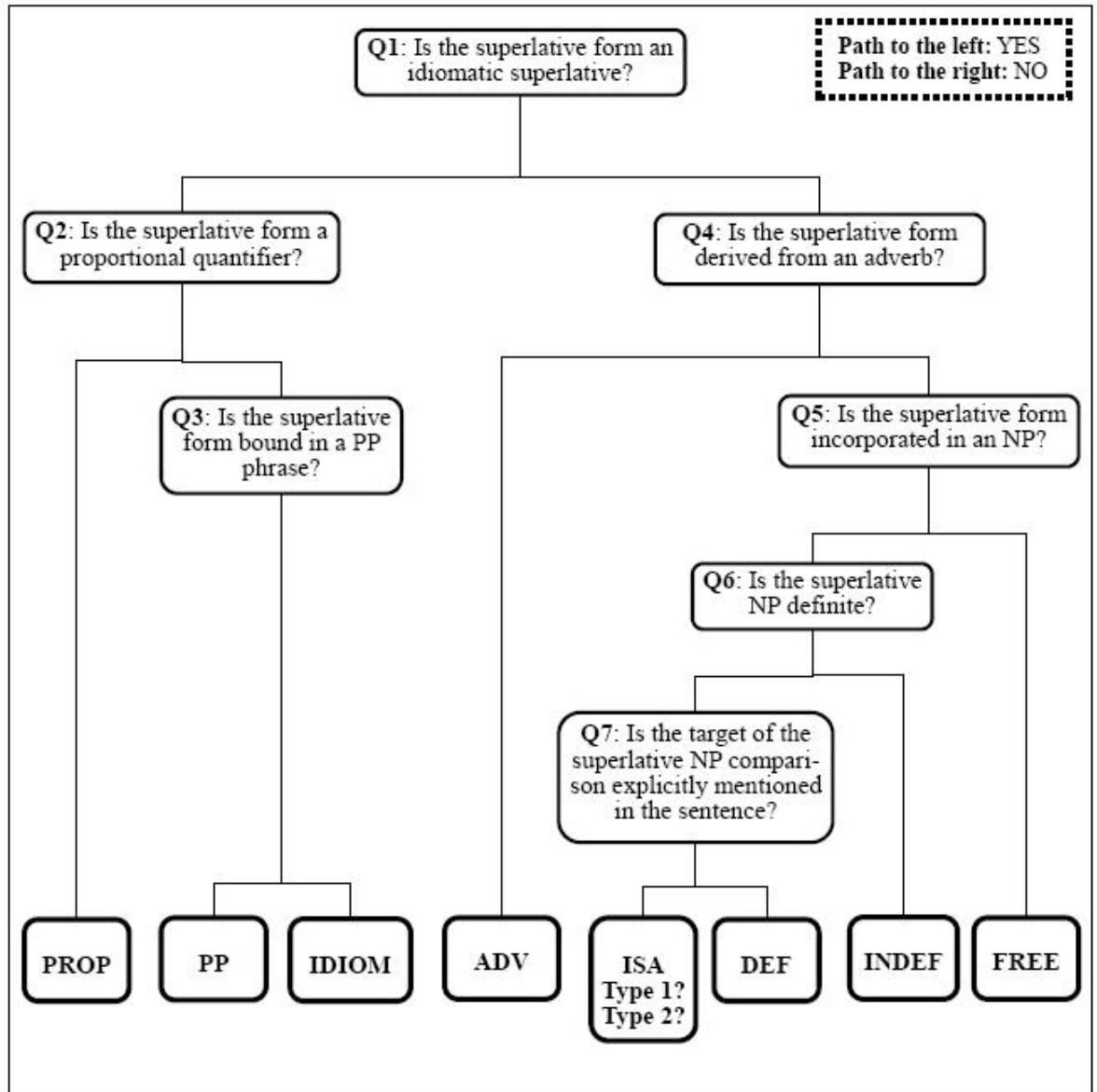


Figure 3.6: Decision tree for superlative classification

each of the points, an arrow refers the reader to the relevant question in the decision tree, and to guidance in the form of definitions which are listed in Appendix B.1:

- *The superlative is **not idiomatic***  
⇒ [Q1: no]; [Guidance: Definition D.1]
- *The superlative is **derived from an adjective***  
⇒ [Q4: no]; [Guidance: Definition D.2]
- *The superlative is **bound in an NP** (including “fused head” constructions)*  
⇒ [Q5: yes]; [Guidance: Definition D.3]
- *The superlative NP is **definite***  
⇒ [Q6: yes]; [Guidance: Definition D.4]
- *The target corresponding to the comparison set expressed by the superlative NP is **explicitly mentioned in the context***  
⇒ [Q7: yes]; [Guidance: cf. “explicit surface form” in 3.1.1]

The class contains the following subgroups:

- **Type 1** (“ISA-1”), which contains surface forms where the target is explicitly mentioned in the same sentence, and the IS-A relation is indicated by the verb “to be” or by apposition. The comparison set may occur either explicitly (as in 3.37 and 3.38) or as a fused head (3.39).
- **Type 2** (“ISA-2”), which contains surface forms where the target and comparison set are explicitly mentioned in the context, and the IS-A relation is indicated by constructions other than the verb “to be” or apposition; e.g. other copula verbs, and constructions such as “verb + *to be*”, “*be* + adjective + *of being/to be*”, or “verb + *as*”.

### Examples

(3.37) {The whale shark} is the *largest* [fish]. [“to be”]

(3.38) The *largest* [fish], {the whale shark}, weighs up to 13.6 tonnes. [*apposition*]

(3.39) (There are many [fish].) The *largest*, {the whale shark}, weighs up to 13.6 tonnes. [*apposition; fused head*]

(3.40) {The whale shark} is considered the *largest* [fish]. [*ISA Type 2*]

## Discussion

The ISA class is very suitable for superlative relation extraction, because its instances *always* express explicit property set comparisons (cf. Section 3.1.1), with both target and comparison set mentioned in the context. This means that all superlatives classified as ISA allow two types of relation to be extracted: 1.) The *IS-A relation* that holds between target and comparison set, and 2.) the *superlative relation* indicated by the superlative form.

**Ambiguities:** In addition to an explicit property set comparison, a superlative classified as ISA may also participate in a **relative set comparison** (cf. Section 3.1.2), as the following example illustrates:

(3.41) Of all the [pupils in Year 7], {Katy} owns the *most expensive* [mobile phone], a {Nokia N95}.

As discussed in 3.1.2, the relative set comparison introduces another level to the property set comparison and restricts the “property” CS members to those who stand in a “verb” relation with the members of the “relative” CS. In particular, this seems to apply to cases where the target of the property set comparison occurs as apposition to a superlative NP (the CS) in object position. In an automated approach which aims to extract superlative targets and comparison sets, this kind of restriction will require special attention, as the extracted property CS string on its own is unlikely to reflect the fact that it is further restricted by a “verb” relation with a second set of entities.

Finally, (3.31) shows that ISA superlatives can also be used in an intensifying way. However, as discussed in 3.1.4, such examples still involve a set comparison, and therefore this does not represent a problem.

**ISA targets and comparison sets:** Another issue concerns the types of targets and comparison sets the ISA class may contain. Most text book examples on superlatives involve comparisons between well-defined *entities* which are designated by noun phrases, which usually have a real-world referent (where the target is a token rather than a type, and the comparison set is a set of tokens rather than a set of types), as in this example:

(3.42) {Mount Everest} is the *highest* [mountain in the world].

In practice, this is not always the case. Firstly, there may be targets and comparison sets that do not have a real-world referent, for instance *classes* of entities (e.g. “{The Border Collie} is the *most intelligent* [dog breed]”). Secondly, while the comparison sets of ISA superlatives are obligatorily noun phrases, with the superlative acting as a premodifier to an NP, targets can also be realised by other phrases:

(3.43) The *best* [approach] is {to wait and see}.

Here, the comparison set consists of different kinds of *approaches*, and the target (i.e. the *best* of these approaches) is expressed by the infinitival phrase *to wait and see*, which is not a single well-defined entity but an abstract action. This is a direct consequence of the fact that the class described by the NP *approach* is itself an abstract noun, which (unlike a word like *dog*) does not represent a well-defined group of entities or classes of entities. Despite having a non-NP target, the classification treats cases like (3.43) as regular members of the ISA class, as it could be substituted by an NP.

**”Fused head” cases:** A further problem concerning membership in the ISA class is represented by the so-called “fused head” constructions (cf. Section 1.3.3), as shown in the following two examples:

(3.44) The Great Dane is the *tallest* of all dogs.

(3.45) The blue whale is the *largest*.

Bos and Nissim (2006) classify such superlatives as “predicative” and exclude them from their considerations (cf. Section 2.2.2). In contrast, I argue that both should be included in the ISA class, as they are property set comparisons with explicit targets and explicit comparison sets. The target strings are easily identifiable: In (3.44), it is *the Great Dane*, while in (3.45) it is *the blue whale*. The difference between the two examples is that in (3.44) the comparison set is also explicitly mentioned in the same sentence, whereas in (3.45) it is not. In fused head constructions, comparison sets are often found in “*of*” prepositional phrases that are either attached to the fused head as postmodifiers, or occur freely in the sentence (comparable to their occurrence in cases like (3.11)). However, they can also be more difficult to identify, and without context even ambiguous. When interpreting (3.45) out of context, one would most likely conclude that the comparison set is the species of whales. However, if the sentence occurred in midst of a discussion about mammals, it would be more likely that the blue whale is compared against the set of mammals. Note that although the comparison set in (3.45) is not explicitly mentioned in the same sentence, it must still be somewhere

in the neighbouring context for the comparison to make sense, which motivates the inclusion of “fused head” cases in the ISA class.

### 3.2.2.2 DEF (Implicit superlative NP comparisons, definite)

#### Surface forms

The DEF class contains definite superlative NP comparisons where the comparison set is explicitly mentioned in the text, but where the target is implicit (i.e. there is no target designator in the sentence). DEF membership is determined by the following criteria:

- *The superlative is **not idiomatic***  
⇒ [Q1: no]; [Guidance: Definition D.1]
- *The superlative is **derived from an adjective***  
⇒ [Q4: no]; [Guidance: Definition D.2]
- *The superlative is **bound in an NP** (including “fused head” constructions)*  
⇒ [Q5: yes]; [Guidance: Definition D.3]
- *The superlative NP is **definite***  
⇒ [Q6: yes]; [Guidance: Definition D.4]
- *The target corresponding to the comparison set expressed by the superlative NP is **not explicitly mentioned** in the context (distinguishes DEF from ISA)*  
⇒ [Q7: no]; [Guidance: cf. “implicit surface form” in 3.1.1]

#### Examples

(3.46) The *largest* [mammal] weighs around 200 tons.

(3.47) Of all the [pupils in Year 7], {Katy} owns the *most expensive* [mobile phone].

(3.48) Beth tells the *best* jokes [when she is drunk].

(3.49) Jim and Julie chose the *most beautiful* music.

(3.50) Tina ordered the *biggest* pizza.

## Discussion

Members of the DEF class can express the following semantic types:

1. *Implicit property set comparisons*, where the superlative NP is used as a referential phrase denoting the target item, as in (3.46)
2. *Relative set comparisons*, where target T2 is left implicit (3.47)
3. *Subject-based set comparisons* involving superlative NPs, as in (3.48)
4. *Intensifying readings* (3.49)

As surface forms like (3.50) (where no CS1 is mentioned) are ambiguous between an implicit property set comparison reading and a relative set comparison reading, the DEF class is of particular interest for theoretical studies of the ambiguity between "absolute" and "relative" readings of superlatives (cf. Chapter 1.3.4).

Syntactically, the members of the DEF class are easy to identify, as the superlative form is incorporated into an NP, and the main verb is usually a non-copula verb. However, there are also cases where the main verb is a copula, as in:

(3.51) Britain's *highest* mountain is in Scotland.

(3.52) The *most beautiful* church in St.Petersburg is on fire.

Both of these are included in the DEF class. This shows that the presence of the verb "to be" is no reliable indicator on its own for inclusion in the ISA class.

### 3.2.2.3 INDEF (Implicit superlative NP comparisons, indefinite)

#### Surface forms

The INDEF class contains cases of NP-bound superlatives which are preceded by an indefinite determiner. INDEF membership is determined by the following criteria:

- *The superlative is not idiomatic*  
⇒ [Q1: no]; [Guidance: Definition D.1]
- *The superlative is derived from an adjective*  
⇒ [Q4: no]; [Guidance: Definition D.2]

- The superlative is **bound in an NP** (including “fused head” constructions)  
⇒ [Q5: yes]; [Guidance: Definition D.3]
- The superlative NP is **indefinite** (distinguishes INDEF from DEF and ISA)  
⇒ [Q6: no]; [Guidance: Definition D.4]

### Examples

(3.53) It’s a *most interesting* book.

(3.54) Coffee has been used as a stimulant from *earliest* times.

(3.55) Tiger, *largest* member of the cat family.

### Discussion

Members of the INDEF class are often used as intensifiers, especially when they have the indefinite article “a” as determiner, as illustrated in (3.53). While (3.54) could be said to also have an intensifying reading, it still seems to involve a set comparison (between different times). Example (3.55), on the other hand, clearly describes a property set comparison, and cases like this often occur in headlines or textual descriptions that aim to provide the reader with an overview of a specific topic. Sentence (3.55) could for example occur as the first sentence in a encyclopedia entry on tigers.

#### 3.2.2.4 FREE

### Surface forms

The FREE class contains adjectival superlatives that are not bound in an NP. In particular, membership in the FREE class is determined by the following criteria:

- The superlative is **not idiomatic**  
⇒ [Q1: no]; [Guidance: Definition D.1]
- The superlative is **derived from an adjective**  
⇒ [Q4: no]; [Guidance: Definition D.2]
- The superlative is **NOT bound in an NP** (distinguishes FREE from INDEF, DEF and ISA)

⇒ [Q5: no]; [Guidance: Definition D.3]

### Examples

(3.56) The African elephant is *tallest* [at the shoulder].

(3.57) *Most common* in Africa is [the zebra finch].

(3.58) The zebra finch is *most common* in Africa.

(3.59) I am *most disappointed* with the decision of the Federal Court.

### Discussion

The FREE class contains superlative instances that can be read as subject-based set comparisons (e.g. (3.56), which compares the height of the African elephant at different locations on its body), set comparisons ((3.57), which compares zebra finches to other finches/birds), and intensifiers (3.59).

As discussed in Section 3.1, ambiguities can arise not only between intensifying and superlative uses, but among the superlative uses there is also frequently ambiguity between subject-based and proper set comparisons. These can only be resolved by investigating the context. However, a closer investigation of such examples showed that the positioning of the sentence constituents may also play a role and could thus help in disambiguating the different readings.

Although the constituents of (3.58) are the same as in (3.57), (3.58) has a different salient reading. This seems to depend on the position of subject and complement in the sentence: In (3.58), involving the normal word order (*Subject (S) - Copula - Complement (C)*), the prominent reading is a subject-based set comparison, where the distribution of the birds in different regions is compared. However, when the positions of subject and complement are swapped round (3.57), the meaning of the sentence is drastically different: What is being compared now is the subject of the sentence (now occurring after the copula) and its complement set (the *zebra finch* is compared to a set of other birds).

### 3.2.3 Adverbial class

#### Surface forms

The ADV class contains superlatives that have been derived from adverbs (Definition D.2). As adverbial superlatives have not been studied much in linguistic literature, they form a class of their own which can be investigated separately in a subsequent step. Adverbial superlatives occur in a variety of syntactic constructions and can modify verbs (3.60 - 3.63), adjectives (3.64), and even whole phrases (3.65).

#### Examples

- (3.60) First Class mail usually arrives the *fastest*.
- (3.61) Tom laughed *loudest* of all.
- (3.62) The lark sings *loudest* when flying fast.
- (3.63) A neutron star radiates *most strongly* at its magnetic poles.
- (3.64) Finland is the *most strongly* pro-European Nordic country.
- (3.65) *Most importantly*, do not forget to turn off the gas.

#### Discussion

Like adjectival superlatives, adverbial ones are generally used for the purpose of comparison and can express a variety of semantic types, similar to the ones described in Section 3.1. Adverbial superlatives can describe set comparisons, where an entity is compared to a comparison set (illustrated by (3.60) and (3.61)), and subject-based set comparisons, where internal states or parts of a single entity are compared (illustrated by (3.62) and (3.63)).

The main difference between adverbial set comparison and adjectival set comparison is that in the former the superlative modifies a verb (or adjective/phrase) and not the set itself. This means that adverbial set comparisons compare a target to its comparison set with respect to *how it performs a particular action*, and not with respect to a general property.

Cases where the adverbial superlative modifies an *adjective* rather than a verb, as for example (3.64), are very similar to the ones described in the ISA class, as the superlative form is incorporated in an NP that represents the comparison set (*Nordic country*) and the target is explicitly mentioned (*Finland*). Although one could argue that cases like this should be included in the ISA class, it seems more useful to keep all adverbial superlatives in a class of their own. The reason for this is that adjectives that can be modified by superlative adverbs in this way are frequently derived from verbs, as in:

(3.66) English is the **most widely spoken** Germanic language.

(3.67) The Internet is the **most quickly developing** mass media in the history of the mankind.

The adverbial superlatives that modify such deverbal adjectives commonly seem to be taken from the class of VP-modifying adverbs, for example adverbs of manner (e.g. *the most quickly accepted technology, the most beautifully situated hotel*), frequency and time adverbs (e.g. *the most regularly/recently updated database*), or adverbs expressing possibility or certainty (*the most likely resulting effect*). This implies that these deverbal adjectives have largely retained their verbal character. Once detached from the nominal head, they can lose their adjectival quality (depending on their state of lexicalisation as an adjective). Consider for example (3.68) and (3.69), which are semantically equivalent to (3.66):

(3.68) Of the Germanic languages, English is (the) *most widely spoken*.

(3.69) Of the Germanic languages, English is spoken *most widely*.

Sentence (3.68) could be interpreted as a “fused head” construction. However, *spoken* could also be reinterpreted as a passive verb form. Likewise, *developing* in (3.67) could be reinterpreted as a progressive verb form once detached from the NP head. Furthermore, the definite article becomes optional. This is in line with Huddleston and Pullum (2002), who describe the difference between incorporated and free superlatives as follows: In free superlatives, the article is always fully optional (cf. Chapter 1.3.3). A special case is represented by cases like (3.70), where the superlatives “most” or “least” are directly followed by the deverbal adjective. Since “spoken” represents an action rather than a property inherent to the members of the comparison set, such cases will be considered as belonging to the adverbial class.

(3.70) English is the **most spoken** Germanic language.

### 3.2.4 Idiomatic superlatives

An idiomatic superlative is a superlative form that is in some way lexicalised. Idiomatic superlative forms are usually not produced by an act of “active” grading (i.e. by an “on-the-fly” derivation of a superlative form from the base adjective/adverb), but are part of a fixed phrase or expression. The proposed classification distinguishes between three different cases, all described below.

#### 3.2.4.1 IDIOM

##### Surface forms

This class contains superlative forms that still constitute a comparison of some sort, but cannot be considered set comparisons as described above because they are lexicalised, and as a result tend to behave differently from conventionally formed superlative phrases. Examples are:

- fixed names/titles of awards: e.g. *Best Western*, the Oscar for *Best Actress*, ...
- uses of superlative forms as nouns: e.g. do my *best*, the *worst* that could happen, ...
- other lexicalised expressions: last but not *least*, the *latest* results, ...

To help in deciding whether a superlative form belongs to the IDIOM class, a “replacement test” can be carried out to assess whether the superlative form can easily be replaced with other superlative forms. For example:

- “*Best Western*” ⇒ “\**Great/Nicest/... Western*” (none of these hotels exist, so *Best* is part of the name)
- “The Oscar for *Best Actress*” ⇒ “The Oscar for \**Greatest/Nicest Actress*” (none of these “Oscars” exist, so *Best* is part of the name)
- “do my *best*” ⇒ “do my \**greatest/nicest*”
- “the *worst* that could happen” ⇒ “the \**most awful/most terrible/... (thing) that could happen*” (superlative form can’t be replaced unless a head noun such as *thing* is included in the sentence)

**Examples**

- (3.71) I'll try my *best*.
- (3.72) Jodie Foster was voted *best* actress in this year's competition.
- (3.73) Last but not *least*, you should always keep your password to yourself.
- (3.74) There are posters up in the Meadows for the *greatest* show on earth.

**Discussion**

As members of the IDIOM class are often highly lexicalised, it is difficult to make general statements about the types of comparisons that they can express. As idiomatic superlatives will not be of interest to the proposed extraction task, their semantics will not be discussed any further.

**3.2.4.2 PP****Surface forms**

The class of PP superlatives contains idiomatic superlative forms that are part of a so-called *PP superlative construction* (Corver and Matushansky, 2006). These are usually phrases of the form [preposition + (*the*) + (*very*/...) + superlative], such as *at most*, *at (the very) least*, *at best*, *at worst*, *at the earliest*, *in the slightest*, *to its fullest*.

**Examples**

- (3.75) At *least* 50 people showed up.
- (3.76) At *best*, people will ignore the message.
- (3.77) I will get there at 8 at the *earliest*.
- (3.78) There, you can find nature at its *finest*.

**Discussion**

Like members of the IDIOM class, PP superlatives are highly lexicalised. For a discussion of their semantics, please see the paper by Corver and Matushansky (2006). As PP superlatives are not of interest for the proposed extraction task, and as they are unlikely to act as potential confusors, their semantics will not be discussed any further.

### 3.2.4.3 PROP

#### Surface forms

The superlative form *most* can be used as a proportional quantifier, usually in phrases of the form “*most* + (*of* + (DET)) + (NP)” (brackets indicate optional constituents). Here, *most* does not have a superlative function but means “more than half” or “the majority”. To distinguish the use of *most* as proportional quantifier from its use as a superlative, the following features intend to provide further guidance:

- **Feature 1:** *most* is preceded by definite determiner  
 ⇒ **Do not classify as PROP**  
 (This excludes (3.79) from being a possible PROP member)
- **Feature 2:** *most* is followed by *of* + NP  
 e.g. *most* of the books, *most* of them  
 ⇒ **Classify as PROP**
- **Feature 3:** *most* is followed by an NP, and it is semantically possible to insert [*of* (+determiner)] between *most* and the NP  
 ⇒ **Classify as PROP**  
 (This identifies (3.81) as a PROP member)
- **Feature 4:** *most* is followed by an NP, and it is semantically possible to precede *most* with *the*  
 ⇒ **Do not classify as PROP**  
 (This excludes (3.82) from being a possible PROP member)

#### Examples

(3.79) She drank the *most* coffee of them all.

(3.80) *Most* books have more than 5 pages.

- (3.81) I've read *most* books written by Virginia Woolf. [The only ones I haven't read are *Jacob's Room* and *The Waves*.]
- (3.82) I've read *most* books written by Virginia Woolf. [I have read 7, Sally has read 5, Peter has read 2, and Toby none.]

### Discussion

Superlatives classified as PROP relate to the semantic type described in Section 3.1.5: They do not involve a superlative comparison, but express a proportional quantification. As they are of little interest for superlative relation extraction, their semantics will not be discussed any further.

## 3.3 Summary

The first part of this chapter introduced and discussed five different types of superlative comparison: Property set comparisons, relative set comparisons, subject-based set comparisons, intensifiers, and proportional quantifiers. In property set comparisons, the target and comparison set stand in an “IS-A” relation, and describe a property (such as height, weight, age...) which all members of the comparison set share, but the target has the highest or lowest value or degree of. In relative set comparisons, two set comparisons are involved: An explicit comparison between a target and its comparison set, which are compared with respect to a second, implicit, property set comparison (involving the superlative form). In subject-based set comparisons a property intrinsic to the subject of the sentence is compared with respect to different “states” or different “parts” of that subject. Finally, intensifiers and proportional quantifiers describe two non-comparative uses of superlatives.

The second part of this chapter proposed an exhaustive classification of superlative surface forms, which allows a detailed study of the relationship between the respective surface forms and their semantic readings. Furthermore, the classification groups together superlative surface forms that are suitable for superlative relation extraction. The ISA class is of particular interest, because its instances *always* express explicit property set comparisons with both target and comparison set designators in the context. The classification of surface forms serves as a basis for the annotation of superlatives described in Chapter 4.

# Chapter 4

## A Corpus of Superlatives

### 4.1 Overview

The first part of this chapter (4.2) describes a two-part annotation scheme for superlatives. The first annotation is based on the classification of superlative forms described in Chapter 3.2. The second one only applies to superlatives that express straight-forward comparisons between targets and their comparison sets (ISA class), and addresses the identification of the spans of each target and comparison set, which is of importance for the proposed superlative relation extraction task. The annotation scheme has been tested and evaluated on 500 tokens of superlatives, the results of which are presented in Section 4.3. In addition to providing a platform for investigating superlatives on a larger scale, this chapter also introduces a text-based Wikipedia corpus (“TextWiki”) which is especially suitable for linguistic research, and in which all superlative occurrences have been annotated with class and target and comparison set information (4.4).

## 4.2 Annotation scheme

### 4.2.1 Classification

#### Overview

As superlatives occur in a variety of different surface forms which tend to express different types of comparisons, it is necessary to deal with these structural classes separately. The classification annotation task aims to classify superlatives according to the classes proposed in Chapter 3.2, one of which is the “ISA-1” class, which will be the basis of the T and CS Identification task (cf. 4.2.2).

#### Procedure

Annotators are asked to mark up each occurrence of a superlative form in the corpus as belonging to one of the eight structural classes defined in Chapter 3.2:<sup>1</sup>

1. IDIOM
2. PP
3. PROP
4. ADV
5. FREE
6. INDEF
7. ISA (Type 1 or 2)
8. DEF

For a given superlative sentence, the task is to decide which of these classes the superlative form belongs to. For this purpose, the annotator is provided with the binary tree displayed in Figure 3.6, where each node represents a question about the superlative form at hand. Starting with the question at the root node of the tree, the annotator moves to the left hand child of a node when he answers the question with “yes”, and to the right hand child of the node when he answers the question with “no”, until he or

---

<sup>1</sup>For the ISA class, there is also a distinction to be made between Type 1 and Type 2.

she reaches one of the leaves of the tree. The annotator then labels the superlative form with the class label indicated by the leaf. As guidance, the annotators were provided with an overview similar to the one in Chapter 3.2.

## 4.2.2 Target and Comparison Set Identification

### Overview

The second annotation applies only to superlatives which are classified as “ISA” members in the previous step. For each ISA instance, the strings representing the *target T* and the *comparison set CS* of the superlative are marked up (cf. Bos and Nissim (2006)). This is crucial for the proposed superlative relation extraction system (cf. Experiment 3 in 6.2).

### Procedure

Annotators are asked to identify and mark up the strings representing the target T and the comparison set CS. For example:

- (4.1)     **Sentence:** Philadelphia Zoo is the *oldest* zoo in America.  
           ⇒ **T:** Philadelphia Zoo  
           ⇒ **CS:** zoo in America.

### Guidance

The main challenge for the annotators is the fact that the comparison set can be restricted in a variety of ways, for example by preceding possessives and premodifiers, or by postmodifiers such as PPs or various kinds of clauses. Compare for example:

- (4.2)     {VW} is [Europe’s] *largest* [manufacturer of cars].  
 (4.3)     {VW} is the *largest* [European car manufacturer with this product range].  
           [restrictive]  
 (4.4)     {VW} is the *largest* [car manufacturer in Europe] with an impressive product range. [non-restrictive]

(4.5) [In China], {VW} is the *largest* [car manufacturer].

The phrases *of cars* and *car* in (4.2) and (4.3) both have the role of specifying the type of *manufacturer* that constitutes the comparison set. The phrases *Europe's*, *European* and *in Europe* occur in determinative, premodifying, and postmodifying position, respectively, but all have the role of restricting the set of car manufacturers to the ones in Europe. And finally, the “with” PP phrases in (4.3) and (4.4) both occur in postmodifying position, but differ in that the one in (4.3) is involved in the comparison, while the one in (4.4) is not. In addition, restrictors of the comparison can also occur elsewhere in the sentence, as shown in (4.5). Similarly, the target can also be pre- or postmodified. It is important that all and only modifiers defining the target or comparison set be identified.

The description of noun phrases has a long tradition in linguistics, and is usually based on the idea that constituents are represented by a continuous sequence of words. Descriptions of the English NP have been proposed in all major grammar theories. In the annotation guidelines, I illustrated the task using the descriptive model proposed by Quirk et al. (1985), which describes the noun phrase as a headed phrase, in which the noun phrase head represents the only obligatory element, while the other constituents (which are the determiner, the premodifier, and the postmodifier) are optional. There can be zero or more premodifiers and/or postmodifiers. The presence of the determiner depends on the realisation of the NP head.

The reason why I chose this model to illustrate the task is that according to Quirk et al., it describes the ordering of the NP constituents in terms of how they really occur in language use. It is therefore assumed that NPs can be split up into the sequence of constituents shown in Table 4.1.

NP			
(DETERMINER)	(PREMODIFIER)*	HEAD	(POSTMODIFIER)*
article	ADJP	noun	PP
pronoun	ADVP	pronoun	CL
numeral	NP	<i>one</i>	ADJP
genitive	NP	nominal adjective	ADVP NP

Table 4.1: NP structure according to Quirk et al. (1985)

An important question here is which of these categories superlative forms belong to. The fact that superlative forms are derived from an open word class (adjectives or adverbs) suggests that they are members of the premodifying class. However, their position is usually closer to the determiner than all other premodifiers. In addition, they behave more like determiners with regards to their function: They single out a unique entity of a set as being of higher or lower rank than all the other entities in the set. This draws attention to a weakness of the above structure: The “determiner” category is actually often realised as a whole phrase (“determinative phrase”). Quirk et al. propose three determiner subfunctions, which are mutually exclusive and have a fixed word order: Predeterminer, central determiner and postdeterminer. They can be realised by the following word classes:

- **Predeterminer:** pronouns (exclamatory, universal, quantifying); numerals (multiples, fractions)
- **Central determiner:** article (definite, indefinite); pronouns (demonstrative, possessive, negative, assertive, relative, non-assertive, interrogative), genitive NP (specifying)
- **Postdeterminer:** numeral (cardinal, ordinal), pronoun (quantifying)

In the annotation of comparison set spans, it makes sense to consider the determinative phrase in terms of these three subcategories. The reason for this is that the superlative form can receive its own category, which prevents us from committing it to one of the above classes, and allows it to occur before the postdeterminer, as for example in “the largest two numbers”. Thus, the superlative NP can be considered to consist of the following components:

Superlative NP						
PREDET	DET	POSTDET	SUP	PREMOD	HEAD	POSTMOD
<i>all, both,</i> ...	<i>a,</i> <i>the,</i> <i>my,</i> <i>Tina's,</i> ...	<i>second,</i> <i>two, little,</i> <i>few, ...</i>	<i>largest,</i> <i>smallest,</i> <i>coolest...</i>	<i>nice, blue,</i> ...	<i>house,</i> <i>bridge,</i> <i>donkey,</i> ...	<i>in Europe,</i> <i>that I have</i> <i>ever seen,</i> <i>ever built...</i>

Table 4.2: CS structure

These constituents usually occur in a linear fashion, with all of the elements apart from SUP and HEAD being optional. In the special case of “fused head” constructions (cf. Chapter 3.2.2.1), the SUP element and the HEAD element are merged.

Annotators were asked to mark up the **determinative phrase** of the superlative and the **main comparison set span**, which is defined to consist of premodifiers, the CS head, and any postmodifiers. External modifiers and preposed elements (as the underlined phrase in (4.6) or “In China” in (4.5)) are not taken into account at this stage. Both target and comparison set are assumed to consist of single uninterrupted strings.

(4.6)     [After the World Cup], the *most important* [football competitions] are {the continental championships}, which are organised by each continental confederation and contested between national teams.

With respect to the determinative phrase, annotators were further instructed to mark up the “extended” determinative phrase where it seemed to indicate the target’s position and uniqueness compared to the other members of the comparison set. For example, while “*the two (largest)*” or “*the other (largest)*” imply non-uniqueness of the target(s), the “extended” determinative phrases “*one of the (largest)*” or “*among the (largest)*” imply that the exact position of the target remains unspecified.

Annotators were further asked to mark up any postmodifiers which in their opinion restricted the comparison set (or target). They did not receive explicit instructions, but were provided with the following list of potential postmodifier types:

- PP: prepositional phrase, e.g. *in the country*
- ADVP: adverbial phrase, e.g. *still, yesterday*
- ADJP: adjectival phrase, e.g. *likely, blue*
- CL-NONF: non-finite clause, e.g. *playing guitar, to sell the house*
- CL-REL: restrictive relative clause (non-restrictive ones are excluded), e.g. *(house) that I own*

- CL-SUB: subordinate clause, e.g. `because they were planted in spring`
- NP: noun phrase, e.g. `park`

No further explicit instructions were given to the annotators so as not to bias their decision. The only further guidance was that annotators should keep the target string as minimal as possible. That is, for cases like (4.7), where the target “The Academy Awards” is followed by further information in brackets, only the main string (describing “essential” information) should be included in the target span.

(4.7) {The Academy Awards} (also known as “the Oscars”) are [the] *most prominent* [film awards in the United States], providing recognition each year to films, ostensibly based on their artistic merits.

### 4.3 Pilot annotation study

Previous experiments have shown that superlatives are particularly frequent in encyclopedia text (Scheible, 2006). For this reason I decided to use Wikipedia<sup>2</sup> as a knowledge base. Before describing the corpus compilation and annotation in more detail, the following sections describe the results of a pilot annotation study that was carried out prior to the TextWiki compilation.

#### 4.3.1 Data and procedure

The pilot annotation study was carried out on a sample drawn from the Wikipedia XML corpus by Denoyer and Gallinari (2006), with sentence mark-up added by Jijkoun and de Rijke (2006). This pilot corpus consists of 142 articles randomly selected from Part-0 of the Wikipedia XML corpus (excluding articles with less than 50 words and all database structures). The corpus contains 500 tokens of superlatives, with (on average) one superlative per 14 sentences.

---

<sup>2</sup><http://en.wikipedia.org>

Superlative instances were identified and annotated as described in Section 4.2. In addition to myself, a second annotator was recruited and trained to test the validity of the proposed annotation scheme. Errors were discussed and resolved after each set of 100 superlative-containing sentences. All annotations were carried out with a tool specifically designed for the task. The following section summarises the results of this pilot study.

### 4.3.2 Results and discussion

#### Classification task

The classification task (Section 4.2.1) achieved an overall inter-annotator agreement of 89% (444/500 instances). Disagreements were discussed after each set of 100 instances, and were, with the exception of three cases, resolved in favour of Annotator 1 (the author). The feedback sessions caused the performance curve to rise steadily, with 76% agreement after the first 100 cases and 95% agreement after the last set:

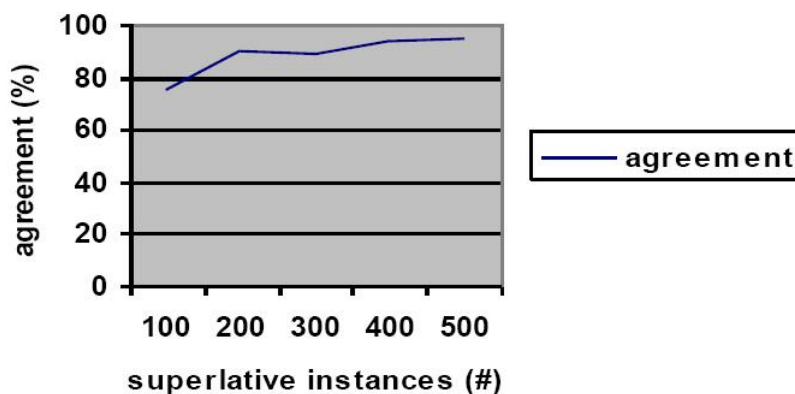


Figure 4.1: Inter-annotator agreement of superlative classification

Surprisingly, the adverbial class ADV was the source of most disagreement, in particular with adverbial superlatives in ISA constructions, as for example in (cf. 3.2.3):

- (4.8) There is not even complete consistency to be found between *The Lord of the Rings* and *The Hobbit*, the two *most* closely related works, because Tolkien was never able to fully integrate all their traditions into each other.

[Article: *J. R. R. Tolkien*]

Despite this structural resemblance to ISA comparisons, these adverbial superlatives do not modify the comparison set head, but rather (one of) its modifiers. In (4.8), *most closely* modifies the deverbal adjective *related* and not the NP head *works*. Another problem was in cases of the form '*superlative + deverbal adjective*', as for example:

- (4.9) An Egyptian scribe named Ahmes wrote the *oldest* known text to give an approximate value for  $\pi$ .

[Pi]

Although similar to *longest-known*, which should be classified as ADV, the superlative *oldest* in this context modifies *text* rather than *known*. (4.9) should therefore be labeled as DEF.

Table 4.3 shows the (agreed) distribution of superlative classes over the 500 superlative sentences.

	<b>TOTAL</b>	<b>%</b>
<b>IDIOM</b>	28	5.6
<b>PP</b>	31	6.2
<b>PROP</b>	124	24.8
<b>ADV</b>	52	10.4
<b>FREE</b>	8	1.6
<b>INDEF</b>	14	2.8
<b>ISA</b>	150	30
- of which <i>ISA-1</i>	108	21.6
- and <i>ISA-2</i>	42	8.4
<b>DEF</b>	93	18.6

Table 4.3: (Agreed) distribution of superlative classes

The ISA class is clearly the most populous, containing 30% of all superlative instances (150/500), which justifies the focus on this class in the second annotation task. The high number of proportional quantifiers (24.8%) can be explained by the fact that encyclopedia entries usually define classes, and proportional quantifiers are a good way of describing properties that do not apply to all members of a class (but to most of

them).

With 99.4% agreed accuracy, Annotator 1's labeling can be reliably used as gold standard for the classification of superlatives in the TextWiki corpus.

### Target and Comparison Set Identification

The results of the Target and Comparison Set Identification for superlatives classified as ISA also look very promising (Section 4.2.2). Of 116 superlatives classified as ISA-1 (89 cases) or ISA-2 (27 cases) by both annotators, there was full agreement for target and comparison set spans in 108 cases (93%). Of the eight disagreements between the two annotators, five concerned the span of ISA-1 targets.

Considering the fact that slightly over two thirds of ISA-1 superlatives (60/89) have a comparison set with at least one postmodifier, the results look very impressive. However, a closer study reveals that only four of these (4/60, around 7%) have a postmodifier that is marked non-restrictive by the annotators. This implies that given a postmodified comparison set, there is a chance of approximately 93% that the postmodifier is restrictive. (The probability may actually be even higher since some comparison sets have more than one postmodifier.)

Compared to ISA-1 comparison sets, ISA-1 targets are less likely to be postmodified: Only 32 out of 89 ISA-1 targets have at least one postmodifier (around 36%). However, the proportion of non-restrictive postmodifiers among them is much higher, with 16 out of 32 ISA-1 targets having a non-restrictive postmodifier (50%). In most cases these are postmodifying clauses (such as relative and non-finite clauses).

According to English comma rules, it should be possible to distinguish between non-restrictive (usually referred to as “non-defining”) and restrictive (“defining”) relative clauses by the presence or absence of a comma. In the following example, the comma after *Ceres* indicates that the following relative clause is to be considered non-defining:

(4.10) [The] *biggest* [asteroid belt member] is {Ceres}, which is about 1000 km across.

[*Asteroid Belt*]

However, commas are not used reliably, as the following sentence illustrates:

- (4.11) [The] *most famous* [diesel-hydraulic locomotive] is {the german V200} which were built from 1953 in a total number of 136.

[*Locomotive*]

If the relative *which*-clause was a defining one, this would imply that there are at least two separate versions of *the german V200*, which is very unlikely.

Interestingly, it seems that distinguishing between defining and non-defining relative clauses also causes problems for the annotators, even when correct comma rules have been applied:

- (4.12) {The temporary exhibition Treasures of Tutankhamun}, held by the British Museum in 1972, was [the] *most successful* [in British history], attracting 1,694,117 visitors.

[*British Museum*]

Annotator 2's target string includes the relative clause "held by...", which should be considered non-restrictive in this context. The unreliability of commas in practice means that annotators must rely on their world knowledge in order to identify target and comparison set spans correctly.

With an inter-annotator agreement of 93%, we can conclude that the identification of target and comparison set spans as defined in 4.2.2 is a fairly straightforward task, and suitable for an automated approach. However, in future work, the following issues should also be taken into account:

- 1.) Around 20% (18 out of 89) of ISA-1 targets or comparison sets contain a pronoun and require anaphora resolution. For example:

- (4.13) [Its] *most populous* [city] is {Vancouver}, which is in the southwest corner of the mainland of the Province of BC.

[*British Columbia*]

- 2.) Around 19% (17 out of 89) of ISA-1 comparison sets contain a "fused head" (cf. Chapter 3.2.2.1). The NP head is implied in the context (usually, but not necessarily in

the same sentence) but has not yet been considered in the annotation, as for example *works* in the following sentence:

- (4.14) He was the author of several works, the *most important* being {The Principles and Practice of Surgery} (1878-1883).

[David Hayes Agnew]

3.) Postmodifiers that have been identified as restrictive should be further analysed according to their semantic roles. In particular, one needs to distinguish between *NP-head complements* and *NP-head modifiers* (Huddleston and Pullum, 2002). In (4.15), the “in” PP-phrase is a complement and an obligatory part of the NP. In (4.16), on the other hand, the “in” PP-phrase is a modifier of the NP head. It has the role of restricting the set in location, but is not as such obligatory:

- (4.15) [The] *newest* [technology in trains] is {magnetic levitation} (maglev).

[Locomotive]

- (4.16) [The] *most popular* [religion in Switzerland] is {Roman Catholicism} (43% of the population).

[Switzerland]

## 4.4 TextWiki corpus

### 4.4.1 Motivation

Existing Wikipedia corpora such as the Wikipedia XML corpus by Denoyer and Gallinari (2006) are primarily aimed at Information Retrieval tasks such as INEX, and have several shortcomings with respect to studying linguistic phenomena like superlatives. Firstly, the XML conversion in Denoyer’s corpus retains most of the original wiki-markup, and thus includes information that is redundant for linguistic investigations (e.g. formatting). Furthermore, database structures such as tables, lists, figures, galleries, and templates are included, which usually do not contain full sentences and may

therefore skew experimental results. Finally, the corpus includes empty or incomplete articles ('stubs') which are of little interest from a linguistic point of view.

I therefore decided to compile a new Wikipedia corpus which is especially suitable for linguistic research (referred to as `TextWiki`). Although marked up in XML, it is primarily text excluding information irrelevant for linguistic investigations:

- All tables, lists, figures, galleries, and templates have been deleted
- Document structure markup is reduced to title, body, paragraph and sentence tags
- Formatting markup is reduced to a small set of tags (e.g. `<b>` for bold text, `<i>` for italics)
- Only hyperlinks to other Wikipedia articles are retained

So far, the `TextWiki` corpus yields over 450,000 words, and all superlative forms in the corpus are marked up and annotated with their class (as described in 4.2.1). In addition, the target and comparison set strings of all ISA superlatives have also been annotated (cf. 4.2.2). Superlatives associated with the words *most* and *least* have been identified, as have adjectives and adverbs graded with the suffix *-est* (including hyphenated superlatives like *kindest-hearted*) and irregular superlative forms (*best*, *worst*, *furthest* and *farthest*).

#### 4.4.2 Structure of the corpus

`TextWiki` aims to be a balanced corpus that draws equally from five different top level Wikipedia categories. Due to Wikipedia's complex categorisation structure and its open-endedness, an exhaustive and even coverage of the whole encyclopedia is problematic. However, the same random sampling technique of articles was applied to each of the five categories, resulting in a corpus containing articles from a broad number of areas. Table 4.4 shows the distribution and total numbers of words (W), articles (A), and superlatives (S) in `TextWiki`.

The table shows around two to three superlatives on average per Wikipedia article.

<b>Top-level Category</b>	<b>W</b>	<b>A</b>	<b>S</b>
Cat1: Culture and the arts	90025	100	291
Cat2: Geography and places	90406	107	360
Cat3: Health and fitness	90174	91	232
Cat4: History and events	90691	98	233
Cat5: Natural and physical sciences	90369	96	264
<b>TOTAL:</b>	451,665	492	1380

Table 4.4: Distribution and total numbers of words (W), articles (A), and superlatives (S) in TextWiki

#### 4.4.3 Distribution of superlative classes in TextWiki

This section describes the results of annotating the 1380 superlatives found in TextWiki with their respective class label. Table 4.5 shows the overall distribution of superlatives classes in the corpus as well as the total number of instances of each class (in brackets).

<b>Class</b>	<b>TextWiki (1380)</b>
<b>IDIOM</b>	3.4% (47)
<b>PP</b>	7.3% (101)
<b>PROP</b>	25.4% (350)
<b>ADV</b>	9.0% (124)
<b>FREE</b>	1.9% (26)
<b>INDEF</b>	1.8% (25)
<b>ISA-1</b>	23.7% (327)
<b>ISA-2</b>	7.3% (101)
<b>DEF</b>	20.2% (279)

Table 4.5: Distribution of superlative classes in TextWiki

The distribution of superlative classes is similar to the one shown in Table 4.3 for the pilot annotation task. The results show that the two most frequent classes are PROP (25.4%) and ISA-1 (23.7%), followed by DEF with 20.2%. The lowest frequency is found for the INDEF and FREE classes with values of 1.8% and 1.9%, respectively. The results lend further support to a focus on the ISA class: combined values for ISA-1 and ISA-2 show that "ISA" is clearly the most frequent superlative type at 31.0% altogether.

#### 4.4.4 Data sets used in the experiments

For the experiments described in the next two chapters, the corpus has been randomly divided up into three parts (by article). Two parts are to be used as development set (DevSet) and development test set (DevTestSet), while the third part serves as test set for the experiments (TestSet). The data sets contain 169, 173, and 149 articles respectively, and the numbers of superlatives in each set (by category, see Table 4.4) are shown in Table 4.6.

<b>Data set</b>	<b>S (Cat1)</b>	<b>S (Cat2)</b>	<b>S (Cat3)</b>	<b>S (Cat4)</b>	<b>S (Cat5)</b>	<b>TOTAL</b>
DevSet	75	121	80	80	78	434
DevTestSet	122	147	86	90	91	536
TestSet	94	93	67	61	95	410

Table 4.6: TextWiki experimental data sets

Table 4.7 shows the distribution of superlative classes across the three data sets.

	<b>DevSet (434)</b>	<b>DevTestSet (536)</b>	<b>Test (410)</b>
<b>IDIOM</b>	1.8% (8)	5.0% (27)	2.9% (12)
<b>PP</b>	9.0% (39)	7.3% (39)	5.6% (23)
<b>PROP</b>	24.7% (107)	23.3% (125)	28.8% (118)
<b>ADV</b>	9.0% (39)	9.3% (50)	8.5% (35)
<b>FREE</b>	3.2% (14)	1.3% (7)	1.2% (5)
<b>INDEF</b>	1.8% (8)	1.9% (10)	1.7% (7)
<b>ISA-1</b>	22.1% (96)	25.9% (139)	22.4% (92)
<b>ISA-2</b>	6.7% (29)	6.9% (37)	8.5% (35)
<b>DEF</b>	21.7% (94)	19.0% (102)	20.2% (83)

Table 4.7: Distribution of superlative classes in TextWiki development and test sets

## 4.5 Conclusion

This chapter described an annotation scheme for superlatives which forms the basis of the experiments described in Chapters 5 and 6. The first annotation deals with superlative classification, while the second addresses the identification of the target

and comparison set spans of superlatives classified as ISA in the previous task. The annotation scheme was tested and evaluated on 500 tokens of superlatives with high inter-annotator agreement for both tasks. A single annotator can therefore be reliably used for gold-standard annotation. Finally, this chapter also introduced a text-based Wikipedia corpus (“TextWiki”) which is especially suitable for linguistic research. All superlative occurrences in this corpus have been annotated with class and target and comparison set information, which will form the basis of the experiments described in the next two chapters.

## Chapter 5

# Identifying and Classifying Superlatives Automatically

### 5.1 Overview

This chapter describes two experiments. The first one aims to automatically identify and mark up superlatives in free text (Section 5.2). The resulting system is called “SUP-Finder” and represents a crucial step in any computational treatment of superlatives, as all other programs rely on the accuracy of this system. The second experiment attempts to classify superlatives according to the classification proposed in Chapter 4.2.1, resulting in a program called “SUP-Classifier” (5.3).

### 5.2 Experiment 1: Identifying superlatives (*SUP-Finder*)

#### 5.2.1 Overview

The first step in a computational treatment of superlatives has to be their identification. The aim of Experiment 1 is therefore to automatically identify and mark up all superlatives in a given text.

As described in Chapter 1.3.1, superlatives are derived from their base adjective/adverb in two different ways: inflectionally or analytically. In the first case, the inflectional suffix *-est* is appended to the base form of the adjective or adverb, while in the second

case they are preceded by the analytical markers *most/least*. In addition, there is a (limited) number of irregular superlative forms, which are listed in Tables 1.1 and 1.2.

Chapter 4.4.1 describes how a pattern-based method was employed to identify all superlatives in the TextWiki corpus. Due to the regularity of the inflectional and analytical patterns and the limited number of irregular cases, it is possible to capture all members of the superlative class by a simple search for the pattern *\*est*, the words *most* and *least*, as well as the irregular cases listed in Tables 1.1 and 1.2. However, this technique requires a substantial amount of manual post-processing, as a large proportion of the retrieved tokens are not superlatives (e.g. “*test*”, “*interest*”). Experiment 1 aims to fully automate the process of superlative identification, which is a crucial step in a computational treatment of superlatives, as all other experiments described in this chapter are based on the output of this system.

## 5.2.2 Method

Previous automatic approaches to identifying superlatives have mainly focussed on techniques involving a search for the POS tags JJS and RBS (e.g. Bos and Nissim (2006)), usually without carrying out a detailed error analysis due to the large amount of manual intervention that would be required for a gold standard. The proposed SUP-Finder aims to improve on the POS-based approach by using a pattern matcher based on regular expressions and a list of “superlative distractors”. In addition, the software TTT2 (Grover and Tobin, 2006) is used to determine sentence boundaries.

The list of superlative distractors was compiled from a BNC word frequency list containing 938,971 different types (downloadable from Adam Kilgarriff’s website<sup>1</sup>). As a first step, all instances matching the pattern *\*est* were identified, yielding 1471 types. Each item of this list was then manually analysed and categorised as either *potential superlative* (768 cases) or *superlative distractor* (703 cases). The decision to have an explicit exclusion (rather than inclusion) list maximises recall, as it allows for very rare superlatives to be included that may not occur in the BNC, as well as neologisms (i.e. new words) and nonce-formations<sup>2</sup>. For this reason, ambiguous and unclear cases were classified as “potential superlative”, and only clear cases of non-superlatives were

---

<sup>1</sup><http://www.kilgarriff.co.uk/bnc-readme.html>

<sup>2</sup>A nonce-formation is “a new complex word coined by a speaker/writer on the spur of the moment to cover some immediate need” (Bauer, 1983).

marked as distractors. Table 5.1 shows examples of both categories.

Potential superlatives	Superlative distractors
<i>best</i>	<i>chest</i>
<i>latest</i>	<i>interest</i>
<i>largest</i>	<i>west</i>
<i>greatest</i>	<i>rest</i>
<i>biggest</i>	<i>test</i>
<i>highest</i>	<i>forest</i>
<i>nearest</i>	<i>suggest</i>
<i>earliest</i>	<i>bequest</i>
<i>lowest</i>	<i>request</i>
<i>finest</i>	<i>honest</i>

Table 5.1: Potential superlatives and superlative distractors

The following bullet points describe the main processing steps of SUP-Finder:

- **Input:** A single text file or directory of text files
- **Step 1:** Determine sentence boundaries with the aid of TTT2 (Grover and Tobin, 2006)
- **Step 2A:** For each sentence, search for the pattern *\*est*, and use the superlative stopword list to exclude non-superlatives
- **Step 2B:** For each sentence, search for the words *most* and *least*, as well as the irregular cases listed in Tables 1.1 and 1.2
- **Step 3:** Wrap the identified instances with XML tags (<superlative>). Note that in cases where the superlative *most* is followed by an adjective, the <superlative> tag only wraps *most*, e.g. <superlative>*most*</superlative> interesting book. The reason for this is that such cases can be ambiguous between proportional quantification (of a noun modified by an adjective) and an analytical superlative (cf. Chapter 3.2.4.3).<sup>3</sup>
- **Output:** An XML file containing all sentences in the given text file(s) that contain a superlative. The superlative form is marked up in XML.

<sup>3</sup>Experiment 2, which deals with superlative classification, aims to resolve this ambiguity.

### 5.2.3 Results

The performance of SUP-Finder is tested on the TextWiki development sets, and compared to a POS-based technique which involves a simple search for the POS tags JJS and RBS (using the output of the C&C tagger (Clark and Curran, 2004b), which is integrated in TTT2). The test corpus contains 7485 sentences in total, 512 of which contain superlatives (i.e. there is around one superlative in every 14-15 sentences). The results are as follows:

Method	Precision	Recall
POS-based Identifier	97.8%	96.1%
SUP-Finder	99.0%	99.8%

Table 5.2: Performance of the superlative identifiers

### 5.2.4 Discussion

Table 5.2 shows that the pattern-based technique outperforms the POS-based one with 99.0% precision and 99.8% recall. The high recall value reflects the fact that superlatives form a well-defined class with a limited number of irregular forms, which makes a pattern-based search fairly straightforward. The POS-based Identifier’s recall, on the other hand, is 3.7% lower than SUP-Finder’s recall, at 96.1%. It also has lower precision values, with 97.8% compared to 99.0% achieved by SUP-Finder.

When evaluating the POS-based approach it is important to bear in mind that the syntactic style of Wikipedia differs from the WSJ, on which the C&C tagger has been trained (even ignoring the spelling and grammatical errors in the former). By not using a tagger trained on the corpus at hand, tagging is likely to be worse than when the tagger is used on the text type it has been developed on. More importantly, tagger performance will have been *optimised* to correctly tag frequently occurring phenomena in its target text type, in order to achieve the highest possible performance score. As superlatives are relatively low frequency phenomena, with most types occurring far down the end of low frequency patterns (part of “the long tail”), even a relatively high-performance tagger like C&C may perform poorly at tagging them, because it will make little difference to the tagger’s overall performance score.

An investigation of the output of the POS-based approach shows that most errors are due to peculiarities concerning the Penn Treebank tagset on the one hand, and tagging errors on the other. For example, hyphenated superlatives such as *best-known* or *longest-living* are categorically tagged as adjectives (JJ) rather than superlatives (JJS). Recall is also affected by tagging errors, which occur frequently in cases where the superlative does not occur in a conventional sentence but in a list (5.1) or a headline (5.2). In both examples, the C&C tagger labelled the superlative instances as proper nouns (NNP).

- (5.1) Positive: \**Longest*<sub>{NNP}</sub> Drive: The longest drive of the group, but it must end up on the fairway.

[*Funnies (golf)*]

- (5.2) *Earliest*<sub>{NNP}</sub> extant maps from the Qin State

[*History of cartography*]

Tagging errors also affect the precision of the POS-based approach, as illustrated by (5.3) and (5.4), where the tokens *lest* and *ingest* were incorrectly tagged as JJS.

- (5.3) In the country however persons engaged in agriculture may freely and lawfully continue their pursuits because it often happens that another day is not suitable for grain-sowing or vine planting; *lest*<sub>{JJS}</sub> by neglecting the proper moment for such operations the bounty of heaven should be lost.

[*Sunday*]

- (5.4) For example, each year millions of seabirds, sea turtles, fish, and marine mammals become entangled in marine debris, or *ingest*<sub>{JJS}</sub> plastics which they have mistaken for food.

[*Debris*]

Both of these examples are included in SUP-FINDER's list of superlative distractors. While SUP-FINDER's recall is nearly 100% (the only error was due to incorrect tokenisation of quotes preceding the superlative form), its precision value of 99% is due to

cases like (5.5), where “MOST” is an acronym, and “asbest” in (5.6), which is missing from the list of superlative distractors.

(5.5) *MOST* was launched in 2003 for the Canadian Space Agency and it is the smallest space telescope in the world, being the size of a small chest or a very large suitcase.

[*Space observatory*]

(5.6) Asbestos : This mineral was extracted in *asbest* mine, near Sverdlovsk from 1889 from the Urals.

[*Siberian natural resources*]

### 5.2.5 Conclusion

With 99.0% precision and 99.8% recall, *SUP-Finder* outperforms the POS-based approach and therefore represents a reliable tool for identifying superlative forms in free text. Unlike the POS-based approach, which has been optimised to work well on a particular text type, the pattern-based approach is independent from text genre and therefore likely to be more robust.

## 5.3 Experiment 2: Classifying superlatives (*SUP-Classifier*)

### 5.3.1 Overview

The second experiment aims to automatically classify superlatives according to the classification proposed in Chapter 3. After experimenting with the development set in Weka,<sup>4</sup> I decided to use a rule-based approach (implemented in Python), using features based on the output of the C&C parser (Clark and Curran, 2004b). This chapter assumes knowledge of common concepts and techniques used in NLP (POS-tagging, lemmatizing, chunking, and NER). For a detailed description of the tools and parser, please refer to (Clark and Curran, 2004b) and the references given below.

---

<sup>4</sup><http://www.cs.waikato.ac.nz/ml/weka/>

### 5.3.2 Tools and approach

*SUP-Classifier* primarily relies on the output of the C&C tools (Clark and Curran, 2004a), which are built around a wide-coverage Combinatory Categorical Grammar (CCG) parser (Clark and Curran, 2004b), and combine various processing components into one single program (“combined analysis”). In addition to the parser, the tools also contain the morphological analyser *morpha* (Minnen et al., 2001), a number of Maximum Entropy taggers, including the CCG supertagger, a POS tagger (Curran and Clark, 2003a), and a chunker and named entity recogniser (Curran and Clark, 2003b). In addition to a grammatical analysis, the tools produce the following levels of annotations, all of which are used as features in the rule-based approach<sup>5</sup>:

1. Lemma
2. POS tag
3. Chunk tag
4. NE tag
5. C&C Supertag<sup>6</sup>

With respect to the C&C parser, the combined analysis tool (“candc”) provides a selection of models and output formats. I decided to use the default output, which is in terms of Briscoe and Carroll style grammatical relations (GR) (Carroll et al., 1999). One of the advantages to this output is that a conversion into a useful format for further processing is fairly straightforward. The output produced by `candc` first lists the set of GR tuples, followed by the lexical annotations for each word in the sentence (using the format `word|lemma|POS tag|chunk|NE tag|supertag`), as shown here for the sentence “The blue whale is the largest mammal”:

```
# this file was generated by the following command(s):
# bin/soap_server --models models --candc-printer xml --server
ltrg.it.usyd.edu.au:9000 --log standard.log

(ncmod _ whale_2 blue_1)
(det whale_2 The_0)
(ncmod _ mammal_6 largest_5)
```

<sup>5</sup>These will be referred to as “lexical annotations”.

<sup>6</sup>The term “supertag” refers to the token’s CCG category, cf. Clark and Curran (2004a).

```
(det mammal_6 the_4)
(xcomp _ is_3 mammal_6)
(ncsubj is_3 whale_2 _)
<c> The|the|DT|I-NP|O|NP[nb]/N blue|blue|JJ|I-NP|O|N/N
whale|whale|NN|I-NP|O|N is|be|VBZ|I-VP|O|(S[dc1]\NP)/NP
the|the|DT|I-NP|O|NP[nb]/N largest|largest|JJS|I-NP|O|N/N
mammal|mammal|NN|I-NP|O|N .|.|.|O|O|.
```

As the `candc` tool expects tokenised sentences as input with one sentence per line, the software TTT2 (Grover and Tobin, 2006) is used for tokenisation. Both software components, TTT2 and C&C, run on Linux machines.

SUP-Classifier has been implemented in Python and consists of two main parts. The first part is a preprocessing pipeline which runs the TTT2 tokeniser and C&C tools on the input file. Table 5.3 shows a brief summary of the main components of this pipeline.

Step	Function	Description
1.	<code>removexml</code>	removes xml markup from input file
2.	<code>tokenise</code>	calls TTT2 to tokenise input text
3.	<code>candcParse</code>	runs <code>candc</code> tools on the input text
4.	<code>processParse</code>	processes <code>candc</code> output file

Table 5.3: Preprocessing pipeline

The second part of SUP-Classifier is the core part where classification is carried out. While the preprocessing pipeline processes the input as a whole, the classification pipeline works on a sentence-by-sentence basis. Initially, for each sentence (5.7), three initial functions `getTags`, `addSuperlativeTag`, and `getParse` retrieve the list of lexical annotations (5.8), add superlative tags to the annotation list (5.9), and retrieve a list of Grammatical Relations (5.10).

(5.7) The largest mammal is the blue whale .

(5.8) `[[The, the, DT, I-NP, O, NP[nb]/N], [largest, largest, JJS, I-NP, O, N/N], [mammal, mammal, NN, I-NP, O, N], ...]`

(5.9) [[The, the, DT, I-NP, 0, NP[nb]/N, 0], [largest, largest, JJS, I-NP, 0, N/N, B-SUP], [mammal, mammal, NN, I-NP, 0, N, 0], ...]  
 (“**TagList**”)

(5.10) [[ncmod, -, 2, 1], [det, 2, 0], [ncmod, 6, 5], ...] (“**GRList**”)

The function `addSuperlativeTag` assigns “Begin/Inside/Outside” (BIO) superlative tags to each token, where the tag B-SUP refers to the superlative start index, and I-SUP refers to any following tokens belonging to the same superlative sequence. In practice, this means that I-SUP marks the predicted superlative end index, as superlatives are taken to be at most two words long. In cases where “most” or “least” is followed by an adjective or adverb (as in “most beautiful”), B-SUP refers to the index of “most” within the sentence, while I-SUP refers to the index of “beautiful”. For inflectional superlatives (e.g. “largest”), B-SUP and I-SUP have the same index, and they are implicitly equated as B-SUP. All other tokens in the sentence receive the “Outside” tag (“O”).

The lists `TagList` and `GRList` are then passed on to the main classification pipeline, which was developed on the `TextWiki` development set and development test set using a cascading approach. This means that classification is carried out in a particular order, where development started out with one class (“IDIOM”), while all other classes were given the label “0”. Once satisfactory precision and recall was reached for the IDIOM class, the next class was added to the class set, which then consisted of IDIOM, PP and 0. In likewise fashion, all other classes were added to the set one by one. Table 5.4 summarises the main steps of the classification pipeline, the last of which (`supClassifier`) is described in detail in the next section.

Step	Module	Description
1.	<code>getTags</code>	gets a list of lexical categories ( <code>TagList</code> )
2.	<code>addSuperlativeTag</code>	adds B-SUP and I-SUP tags to <code>TagList</code>
3.	<code>getParse</code>	gets a list of GR ( <code>GRList</code> )
4.	<code>supClassifier</code>	carries out classification using <code>TagList</code> and <code>GRList</code>

Table 5.4: Classification pipeline

### 5.3.3 Classification strategy: The `supClassifier` module

Classification is carried out by `supClassifier` (Step 4 in the classification pipeline, cf. Table 5.4), which consists of a cascading set of modules, each of which contains a number of diagnostic tests to determine class membership. The cascading architecture means that as soon as one of the modules returns a match, the superlative instance at hand is labelled with the according class label, and the classifier moves on to the next sentence. The order in which classification modules are applied is shown in Table 5.5.

Part I	Part II
1. IDIOM	5. FREE
2. PP	6. INDEF
3. PROP	7. ISA-1
4. ADV	8. ISA-2
	9. DEF

Table 5.5: Order of classification

The first four modules in Part I (IDIOM, PP, PROP, and ADV) mainly use features based on the lexical annotations provided in `TagList`. The modules in Part II, on the other hand, require a substantial amount of syntactic information to carry out classification. For instance, information is needed about whether the superlative form is bound in an NP, and if so, what the NP head is, and whether there is a determiner. To reduce the processing load, `SUP-Classifier` carries out the following tests, whose results are passed on to the classification modules in Part II: First of all, a function `findHead` carries out various tests to check whether the superlative is bound in an NP, and if successful returns the index of the NP head (which will be referred to as “CSHead” in the following sections). If there is a CSHead, a further function attempts to identify a determiner (`findDeterminer`). Finally, for cases where no determiner is found, a third function tests whether the superlative NP is part of a coordinated list (`checkCoordList`). Table 5.6 summarises the main terminology used in the next sections, where all classification modules are described in more detail.

Name	Description	Example
TagList	List of lexical annotations	[[The, the, DT, I-NP, 0, NP[nb]/N, 0], [largest, largest, JJS, I-NP, 0, N/N, B-SUP], [mammal, mammal, NN, I-NP, 0, N, 0], ...]
GRList	List of Grammatical Relations (GR)	[[ncmod, -, 2, 1], [det, 2, 0], [ncmod, 6, 5], ...]
B-SUP	Superlative start index	The index of largest is 1
I-SUP	Superlative end index	Same as B-SUP, 1
CSHead	Index of superlative NP head, if applicable	Index 2 (mammal)
DET	Index of superlative determiner, if applicable	Index 0, (The)

Table 5.6: Terminology (Example: “The largest mammal is the blue whale .”)

### 5.3.3.1 IDIOM

As SUP-Classifier works with a cascading architecture, members of the IDIOM class (see Chapter 3.2.4.1) need to be identified first as their structure often resembles instances of other classes. IDIOM-Identifier employs three main strategies for identifying idiomatic superlatives. The first one, referred to as CapsTest method, checks whether the superlative form is capitalised. Two cases are distinguished: If the superlative start word (in B-SUP position) is not the first word in the sentence, CapsTest succeeds when B-SUP is capitalised, and the superlative is classified as IDIOM, as in (5.11). In cases where B-SUP occurs in sentence-initial position, CapsTest investigates the second word and classifies the superlative as IDIOM only if the second word is also capitalised (5.12). However, cases where B-SUP position is occupied by the word *most* are excluded because they are likely to be proportional quantifiers (PROP) modifying capitalised nouns, as for example in (5.13).

- (5.11) Maddox once pulled an infamous April Fools’ Day joke on April 1, 2004, on his site, The *Best Page In The Universe*.

[April Fools’ Day]

- (5.12) *Longest Drive* and *Nearest the Pin* are most usually competed for by all of those taking part on Golf Society or corporate days with prizes for the winners.

[*Funnies (golf)*]

- (5.13) *Most* Christians today consider Sunday to be the Sabbath day, a holy day and a day of rest and church-attendance.

[*Sunday*]

The second strategy employed by `IDIOM-Identifier` makes use of a superlative idiom lexicon (`IdiomLex`), which contains an extendable list of fixed idioms containing superlatives (such as “last but not least” or “most common multiple”). The superlative instance and its context are compared to each lexicon entry word-by-word, and if an exact match is found, the superlative is labelled as `IDIOM`. Example (5.14) would be recognised as `IDIOM` because the phrase “best practices” is included in `IdiomLex`.

- (5.14) Evidence-based practice (EBP) develops individualized guidelines of *best* practices to inform the improvement of whatever professional task is at hand.

[*Evidence based practice*]

The third method used to identify members of the `IDIOM` class consists of a number of lexical and syntactic patterns, e.g. cases where the superlatives “best” or “worst” are preceded by a possessive pronoun (tag `PRP$`) and not followed by a noun, as for example in (5.15), or superlatives that are preceded by the lemmas “record/award/Oscar” and “for”, as in (5.16). Syntactic patterns include cases like (5.17), which are identified by searching the `GRList` for relations where the superlative “best” or “worst” stands in a `dobj` relation with the lemma “do”, e.g. in (5.17) the relevant tuple is (`dobj do_13 best_9`).

- (5.15) Horonuku was not renowned as a warrior, but was an intelligent and far sighted statesman who did his<sub>{*PRPS*}</sub> *best* for his people.

[*Te Heuheu Tukino IV*]

- (5.16) Gagne went from an average pitcher to being hall of fame eligible, winning the National League Cy Young Award in 2002, by tying the National League record for *most* saves in a season, and the National League Roloids Relief Man of the Year in 2002 and 2003.

[Cyborg]

- (5.17) Since there is no cure for dementia, the *best* an individual can do is to prevent it from developing in the first place.

[Dementia]

Table 5.7 summarises the main strategies of IDIOM-Identifier.

Method	Short description	Examples
I.	CapsTest: capitalisation of B-SUP?	(5.11), (5.12)
II.	IdiomLex: listed in Idiom Lexicon?	(5.14)
III.	Lexical/syntactic pattern?	(5.15), (5.16), (5.17)

Table 5.7: Summary IDIOM-Identifier

### 5.3.3.2 PP

Members of the PP class (described in Chapter 3.2.4.2) are identified via three main rules involving lemmas and POS-tags. The first rule simply states that all instances of the lemma “least” preceded by “at” are classified as PP (5.18). The second rule succeeds if the lemma preceding the superlative in B-SUP position is “at”, and the word following I-SUP does not have a POS-tag NN, NNS, or NNP, or lemma “of”. This matches cases like (5.19), but excludes sentences such as “at *most* (of these) venues smoking is forbidden” or “we will look at *best* prices”, which should be classified as members of the PROP and INDEF classes, respectively.<sup>7</sup>

- (5.18) A chain of stones must have at *least* one liberty to remain on the board.

[Go (board game)]

<sup>7</sup>No such examples occurred in the development data.

- (5.19) However the number of lives lost was surprisingly small; it is believed to have been 16 at *most*.

[Stuart London]

The third PP-identifying rule matches phrases such as “in the slightest” or “at the (very) earliest”. Here, B-SUP has to be preceded by “in/at” + “the” + (“very”), and I-SUP may not be followed by the POS-tags NN, NNS, NNP, JJ, or the lemma “of”, thus identifying (5.20), but excluding cases like (5.21), which should be classified as DEF.

- (5.20) Today there are very few Afro-Chileans, at the *most*, fewer than 1% can be estimated from the 2006 population.

[Afro-Latin American]

- (5.21) In the *rarest* of these sorts of hearing loss, only the auditory centers of the brain are affected.

[Hearing impairment]

Table 5.8 summarises the rules used by PP-Identifier.

Method	Rule	Example
I.	[“at” + “least”]	(5.18)
II.	[“at” + B-SUP] $\wedge$ [I-SUP + $\neg$ {NN, NNS, NNP, “of”}]	(5.19)
III.	[“at”/“in” + “the” + (“very”) + B-SUP] $\wedge$ [I-SUP + $\neg$ {NN, NNS, NNP, JJ, “of”}]	(5.20)

Table 5.8: Summary PP-Identifier

### 5.3.3.3 PROP

Since proportional quantifiers are formed by use of the words “most” and “least” (cf. Chapter 3.2.4.3), PROP-Identifier only considers superlatives whose lemma in B-SUP position equals “most” or “least”. The main challenge in identifying PROP

superlatives is in cases where “most” (or “least”) is followed by an adjective or adverb, as without taking context into account they can be ambiguous between “most” as proportional quantifier and an analytical superlative form involving the following adjective or adverb. Consider for example (5.22), where *most* is a proportional quantifier rather than part of the analytical superlative *most malignant*:

- (5.22) Contrast agent uptake, sometimes in characteristic patterns, can be demonstrated on either CT or MRI-scans in *most malignant* primary and metastatic brain tumors.

[*Brain tumor*]

If the POS-tagging of “most” and “least” was reliable, identification of PROP cases would be straightforward: Instances of “most” or “least” expressing proportional quantification as in (5.22) should be tagged JJS, while the analytical alternative should receive the tag RBS (adverbial modifying the following adjective/adverb). An investigation of “most” + JJ sequences in the development corpus, however, has shown that the output of the POS tagger is unreliable. As previously mentioned, statistical taggers are developed to deal efficiently with phenomena that occur frequently in the text type they are aimed at, but may perform not so well on less frequent phenomena like superlatives. As the tag JJS cannot be used as reliable indicator for identifying PROP instances, the PROP method aims to take the anticipated tagging errors into account.

PROP-Identifier’s strategy consists of three methods. The first two list a number of lemma/POS-based patterns which describe rules of exclusions and inclusions in the PROP class respectively. An example of a rule of exclusion is the sequence “the” + B-SUP (5.23), as the definite article usually indicates a superlative NP (which is classified as DEF or ISA, depending on the presence or absence of a target entity). Also excluded are all cases where the tag of the following token is RB (adverb). These cases usually belong to the ADV class, where the superlative modifies a clause, as for example in (5.24).

- (5.23) The *most* important Siberian petroleum zones are the Central Urals, Sakhalin Island, Nordvyl, on the Arctic Siberian coast and the Kamchatka peninsula.

[*Siberian natural resources*]

- (5.24) *Most* notoriously<sub>{RB}</sub>, the Highland Clearances in northern Scotland led to significant depopulation.

[*Abandoned village*]

A rule of inclusion is “most” (or “least”) followed by the lemmas “of” or “other”, which are always classified as members of the PROP class (5.25). Also included are cases where “most” or “least” carry the supertag N, implying that they are NP heads, as in (5.26), unless followed by a token tagged RB (adverb), in which case they are likely to belong to the ADV class (cf. 5.24 above).

- (5.25) A large number of observatories have been launched into orbit, and *most* of them have greatly enhanced our knowledge of the cosmos.

[*Space observatory*]

- (5.26) While *most*<sub>{N}</sub> were involved in the effort to build the canal, many also came to work on Panama’s banana plantations.

[*Afro-Latin American*]

In the third part, PROP-Identifier deals with cases of “most” and “least” that have not been included or excluded by the rules in Parts I and II. In particular, it aims to deal with the problematic cases that are ambiguous between a PROP reading and one involving an analytical superlative (described above). The Identifier inspects the token following “most” (or “least”), and distinguishes between three different cases:

1. Case 1: The next token has POS tag JJ
2. Case 2: The next token has POS tag VBN or VBD
3. Case 3: The next token is not tagged JJ, VBN or VBD

In the first case, where the next token is tagged as an adjective (JJ), PROP-Identifier first checks the POS tag of the word preceding the superlative, and employs different strategies depending on whether “most” or “least” has been tagged JJS or RBS. For example, if “most” (or “least”) is tagged JJS, and the preceding word is a conjunction (CC) or comma (,), then PROP-Identifier succeeds (5.27). If, however, the superlative has been tagged RBS after a conjunction or comma, PROP-Identifier only suc-

ceeds if the word before the conjunction/comma is a noun (NN, NNS, or NNP), as in (5.28). If none of these rules apply, *PROP-Identifier* makes a decision based on the length of the token tagged JJ. If its length is smaller than 6, the Identifier classifies the instance as member of PROP (5.29), and quits otherwise (5.30). This decision is based on the insight that analytical patterns usually apply to adjectives of at least two syllables (cf. Chapter 1). The threshold of 6 was determined after experimenting with the development sets.

- (5.27) In fact,<sub>{,}</sub> *most*<sub>{JJS}</sub> commercial<sub>{JJ}</sub> navigational maps, such as road maps and town plans, sacrifice an amount of accuracy in scale to deliver a greater visual usefulness to its user, for example by exaggerating the width of roads.

[Map]

- (5.28) Before the advent of fiber optic transmission<sub>{NN}</sub> ,<sub>{,}</sub> *most*<sub>{RBS}</sub> long<sub>{JJ}</sub> distance telephone calls were carried via microwave point-to-point links through sites like the AT&T Long Lines.

[Microwave]

- (5.29) Yet the migration is gradual, and as of 2005 *most* major<sub>{<6}</sub> motion pictures are still recorded on film.

[Film]

- (5.30) The spots are *most* abundant<sub>{>6}</sub> on the thighs and legs, and a person with the ailment looks pale, feels depressed, and is partially immobilized.

[Scurvy]

Cases of “most” and “least” whose next token has the POS-tag VBD or VBN (Case 2 above) are classified as PROP if *B-SUP* is the first word in the sentence, or if it follows a comma or conjunction (tag CC). This classifies cases like (5.31) as PROP, but excludes (5.32), which is a member of the adverbial class (ADV). Although the use of the past participle form of the verb (VBN) indicates ADV membership in (5.32), a general rule that excludes patterns of “most” + VBN from PROP membership is not useful, as

the tagger often fails to distinguish between VBD (verb, past tense) and VBN (verb, past participle), especially in cases where the past participle and past tense forms are homographs. For this reason, cases of “most” and “least” whose next token has the POS-tag VBN are also considered by the above rule, helping to identify incorrectly tagged cases like (5.33) as proportional quantifier.

(5.31) Although they planned to return to Haiti ,{,} *most* stayed<sub>{VBD}</sub> on in Cuba.

[Afro-Latin American]

(5.32) This and Medical Model are probably the ones<sub>{NNS}</sub> *most*<sub>{RBS}</sub> used<sub>{VBN}</sub> by non-disabled people to define and explain disability.

[Disability]

(5.33) *Most*<sub>{JJS}</sub> used<sub>{VBN}</sub> Ptolemy’s methods; but they also took advantage of what explorers and merchants learned in their travels across the Muslim world, from Spain to India to Africa, and beyond in trade relationships with China, and Russia.

[History of cartography]

Finally, if the token following B-SUP is not tagged JJ, VBN or VBD (Case 3.), PROP-Identifier fails. Table 5.9 gives a summary.

Method	Short description	Example
I.	Rules of exclusion	(5.23), (5.24)
II.	Rules of inclusion	(5.25), (5.26)
III.	Strategies for “most”/“least” + JJ or VBD/VBN	(5.27), (5.28), (5.29), (5.31), (5.33)

Table 5.9: Summary PROP-Identifier

#### 5.3.3.4 ADV

The Identifier for the ADV class (which contains adverbial superlatives, cf. Chapter 3.2.3) has four parts altogether. In the first part, ADV-Identifier employs a number

of inclusion tests, for example checking for combinations of “most” or “least” followed by an adverb (RB), as in example (5.24). Furthermore, superlatives that have an I-VP chunk in B-SUP position are also classified as members of ADV (5.34).

- (5.34) A famous example of a map without scale is the London Underground map, which *best*<sub>{I-VP}</sub> fulfils its purpose by being less physically accurate and more visually communicative to the hurried glance of the commuter.

[Map]

Next, ADV-Identifier carries out a “compound search” for cases like (5.35), where the superlative marker is followed by a hyphen. In such constructions, the superlative form usually modifies a verb or deverbal adjective tagged VBG or VBN, and should therefore be classified as ADV. As such cases can also occur without a hyphen, as for example in (5.36), one might consider carrying out a simple search for combinations of *superlative* + VBG/VBN to capture unhyphenated cases as well. However, this is likely to decrease precision, as the superlative does not necessarily modify the deverbal adjective, as illustrated by (5.37). Here, the superlative *oldest* modifies the NP head *map* and not the deverbal adjective *surviving*.

- (5.35) After Spanish, Creole is the second *most*-spoken language in Cuba.

[Afro-Latin American]

- (5.36) The population was 193,830 in 1999 (census), and it is currently the *fastest* growing town in Kenya, and currently the 5th largest in Kenya.

[Eldoret]

- (5.37) It is the *oldest* surviving world map from East Asia, and the oldest Asian map to depict the Western world

[Ancient world maps]

As cases like (5.36) and (5.37) strongly depend on the semantics of the superlative and the following deverbal adjective, the third method consists of a number of lemma-based searches, such as *longest* + *lasting/living/running*, or *fastest* + VBG, which can

identify cases like (5.36).

ADV-Identifier’s last method makes use of the GR annotation and addresses cases where superlatives modify verbs. For example, in (5.38), the superlative modifies the verb *know*, and the word in I-SUP position (*likely*) stands in an xcomp relation via the complementiser “to” (xcomp to\_24 likely\_23 know\_25):

- (5.38) One of the key features of Milgram’s methodology is that participants are asked to choose the person they know who is *most* likely to know the target individual.

[*Small world experiment*]

Table 5.10 provides a summary of the main methods.

Method	Short description	Example
I.	Inclusion tests	(5.24), (5.34)
II.	Compound search: I-SUP + hyphen	(5.35)
III.	Lemma-based searches	(5.36)
IV.	Syntactic patterns	(5.38)

Table 5.10: Summary ADV-Identifier

### 5.3.3.5 FREE

Superlatives classified as FREE are not bound in a noun phrase but occur freely in the sentence (cf. Chapter 3.2.2.4). FREE-Identifier therefore makes use of the results of the two modules described at the beginning of the current section, `findHead` and `findDeterminer`, and consists of two rules: The first one checks if `findHead` has identified a head. If no head has been found, the Identifier classifies the superlative as member of the FREE class:

- (5.39) If more than one person succeeds, the funny goes to the one who was *nearest* the flag.

[*Funnies (golf)*]

The second rule deals with cases where `findHead` identified a fused head, i.e. cases where the head coincides with the word in I-SUP position. In such cases, `FREE-Identifier` checks the results of `findDeterminer`. If no determiner has been identified, the superlative is classified as FREE (5.40).

(5.40) Active, gestural, or challenging standing poses are often scheduled at the beginning of a session when the models' energy level is *highest*.

[*Model (art)*]

Table 5.11 summarises `FREE-Identifier`'s main methods.

Method	Short description	Example
I.	Module <code>findHead</code> does not find a <code>CSHead</code>	(5.39)
II.	Module <code>findHead</code> identifies a fused <code>CSHead</code> ; Module <code>findDeterminer</code> does not find determiner	(5.40)

Table 5.11: Summary `FREE-Identifier`

### 5.3.3.6 INDEF

Members of the `INDEF` class are characterised by being bound in an NP but lacking a definite determiner (cf. Chapter 3.2.2.3). `INDEF-Identifier` therefore uses two main strategies: The first one checks the results of the modules `findHead` and `findDeterminer`, and classifies the instance as `INDEF` if the former succeeds but the latter fails (5.41).

(5.41) He was successful in business, and became president of the American Net and Twine Company, *largest* manufacturer<sub>{`CSHead`}</sub> of its kind in the world at the time.

[*Ivers Whitney Adams*]

If there is a determiner, `INDEF-Identifier` carries out a second test to check whether the determiner is indefinite (“a” or “an”). If this is the case, the instance is also labelled `INDEF` (5.42).

(5.42) It's a *most* interesting book.

(Note that although the roles of the superlatives in (5.41) and (5.42) differ greatly from a semantic point of view, they are included in the same syntactic class (cf. discussion in Chapter 3.2.2.3).)

A summary is provided in Table 5.12.

Method	Short description	Example
I.	Module <code>findHead</code> succeeds, but <code>findDeterminer</code> fails	(5.41)
II.	Modules <code>findHead</code> and <code>findDeterminer</code> succeed, and the determiner is indefinite (“a” or “an”)	(5.42)

Table 5.12: Summary INDEF-Identifier

### 5.3.3.7 ISA-1

ISA-1 superlatives are bound in a definite noun phrase, and the target of comparison is explicitly mentioned in the sentence (Chapter 3.2.2.1). Due to this dependency on a target, `ISA1-Identifier` makes extensive use of the Grammatical Relations output of the C&C parser. Once it has established that there is a `CSHead` (checking the results of `findHead`) and a determiner (via `findDeterminer`), two main cases are distinguished: Instances where the IS-A relation between target and comparison set is expressed via the verb “to be”, and cases where the relation is expressed via apposition. The strategy of the Identifier for identifying the former is as follows:

- Step 1: Locate the position of `CSHead` within the sentence (via subject relation “`ncsubj`” or complement “`xcomp`”)
- Step 2: Test whether the relation word between the `CSHead` and its dependant is a form of “to be”
- Step 3: Find the corresponding target entity

If all three steps succeed, the instance is classified as ISA-1.

The Identifier addresses the first step by testing whether the head of the superlative NP (`CSHead`) occurs in `ncsubj` position, as for example in (5.43).

(5.43) The *largest* mammal<sub>{CSHead}</sub> is the blue whale.

The output of the C&C parser for this example is displayed in Table 5.13. To fulfil Step 1 above, the Identifier first searches for a GR tuple where CSHead (here: *mammal*) stands in an *ncsubj* position (Row 6 in Table 5.13). Typically, first slot in the tuple is occupied by the string “ncsubj”, and the third one is occupied by the CSHead *mammal*. Step 2 is then met by checking if the item in the second slot is a form of “to be”. If it is, Step 3 is addressed by searching the GR list for a tuple where the identified verb stands in an *xcomp* relation with another word (the suspected target), as shown in Row 5.

Row	GR output	Notes
1	(ncmod _ mammal_2 largest_1)	
2	(det mammal_2 The_0)	
3	(ncmod _ whale_6 blue_5)	
4	(det whale_6 the_4)	
5	(xcomp _ is_3 <b>whale</b> _6)	<b>Target</b> is in <i>xcomp</i> position via “is”
6	(ncsubj is_3 <b>mammal</b> _2 _)	<b>CSHead</b> is in <i>ncsubj</i> position via “is”

Table 5.13: GR output for “*The largest mammal is the blue whale.*”

Cases where the relation between CSHead and target is reversed are dealt with in like-wise manner (5.44). The relations are shown in Table 5.14.

(5.44) The blue whale is the *largest* mammal.

Instances where the ISA relation is expressed via apposition receive a different treatment, and the following general steps are applied:

- Step 1: Test whether CSHead stands in a “conj” relation with anything (but excluding instances of “and”)
- Step 2: Search the GR List for the “linked” item (the suspected target)
- Step 3: Locate the position of the CSHead and the target in the sentence (“ncsubj” or “dobj”)

Row	GR output	Notes
1	(ncmod _ whale_2 blue_1)	
2	(det whale_2 The_0)	
3	(ncmod _ mammal_6 largest_5)	
4	(det mammal_6 the_4)	
5	(xcomp _ is_3 <b>mammal_6</b> )	<b>CSHead</b> is in xcomp position via “is”
6	(ncsubj is_3 <b>whale_2</b> _)	<b>Target</b> is in ncsbj position via “is”

Table 5.14: GR output for “*The blue whale is the largest mammal.*”

First, `GRList` is searched for tuples where `CSHead` stands in a `conj` relation with another word. For example, in the following sentence, Step 1 identifies the comma with index 3 as potential appositive conjunction (cf. Table 5.15, Row 5):

(5.45) The blue whale, the *largest* mammal, eats krill.

Addressing Step 2, the Identifier then searches for another tuple in `GRList` with “`conj`” in first position and the comma with index 3 in second position (finding the tuple in Row 6). This is identified as a potential “linked” target. Step 3 distinguishes cases like (5.45), which appear in subject position, from cases like (5.46) below, which appear in object position. A test is carried out to determine whether both target and `CSHead` stand in a `ncsubj` position via the same word. Rows 8 and 9 in Table 5.15 show that both *mammal* and *whale* stand in the required relation via the word “eats” with index 8. The Identifier therefore concludes that the superlative appears in an apposition and classifies it as ISA-1.<sup>8</sup>

Appositions in `dobj` position are treated in a similar manner. For an example like (5.46), `ISA1-Identifier` determines that both target and `CSHead` are in a `conj` position via the same comma (cf. Rows 6 and 7 in Table 5.16), and that they stand in a `dobj` position via the same word “hunted” (cf. Rows 8 and 9). The Identifier concludes that the superlative appears in an apposition and classifies it as ISA-1.

<sup>8</sup>Similar to (5.43) and (5.44), it does not make a difference which item occurs within the apposition. This means that the Identifier also succeeds for the sentence “The largest mammal, the blue whale, eats krill.”

Row	GR output	Notes
1	(ncmod _ whale_2 blue_1)	
2	(det whale_2 The_0)	
3	(ncmod _ mammal_6 largest_5)	
4	(det mammal_6 the_4)	
5	(conj ,_3 <b>mammal_6</b> )	<b>CSHead</b> is in a “conj” position via a comma
6	(conj ,_3 <b>whale_2</b> )	<b>Target</b> is in a “conj” position via the same comma
7	(dobj eats_8 krill_9)	
8	(ncsubj eats_8 <b>mammal_6</b> _)	<b>CSHead</b> is in “ncsubj” position via “eats”
9	(ncsubj eats_8 <b>whale_2</b> _)	<b>Target</b> is in “ncsubj” position via the same word, “eats”

Table 5.15: GR output for “*The blue whale, the largest mammal, eats krill.*”

(5.46) Whaling nations have hunted the blue whale, the *largest* mammal, close to extinction.

In addition to the methods described above, ISA1-Identifier also employs a number of minor strategies to help identify ISA-1 instances involving special constructions like “of which” (5.47) and “among” (5.48).

(5.47) There are many different ocean animals, the *largest* of which is the blue whale.

(5.48) The blue whale is among the *most* endangered marine life species.

Table 5.17 gives a summary of the main strategies of ISA1-Identifier.

Row	GR output	Notes
1	(ncmod _ nations_1 Whaling_0)	
2	(ncmod _ whale_6 blue_5)	
3	(det whale_6 the_4)	
4	(ncmod _ mammal_10 largest_9)	
5	(det mammal_10 the_8)	
6	(conj ,_7 <b>mammal_10</b> )	<b>CSHead</b> is in a “conj” position via a comma
7	(conj ,_7 <b>whale_6</b> )	<b>Target</b> is in a “conj” position via the same comma
8	(dobj hunted_3 mammal_10)	<b>CSHead</b> is in “dobj” position via “hunted”
9	(dobj hunted_3 whale_6)	<b>Target</b> is in “dobj” position via the same word, “hunted”
10	(dobj to_13 extinction_14)	
11	(iobj close_12 to_13)	
12	(ncmod _ hunted_3 close_12)	
13	(aux hunted_3 have_2)	
14	(ncsubj hunted_3 nations_1 _)	

Table 5.16: GR output for “*Whaling nations have hunted the blue whale, the largest mammal, close to extinction.*”

### 5.3.3.8 ISA-2

The ISA-2 class contains surface forms where target and comparison set are explicitly mentioned in the context, and the IS-A relation is indicated by constructions other than the verb “to be” or apposition (Chapter 3.2.2.1). As this involves a variety of syntactic constructions, it is not possible to find one general method that is able to capture all instances. Instead, *ISA2-Identifier* makes use of a number of syntactic rules, each of which has its own lexicon of associated ISA-2 indicators. These rules are based on the data encountered in the development sets, and on intuition. Nine main ISA-2 patterns are distinguished altogether, illustrated by the following examples, where the

Method	Short description	Example
I a.	CSHead is in <i>nsubj</i> position via the verb “to be”	(5.43)
b.	CSHead is in <i>xcomp</i> position via the verb “to be”	(5.44)
II a.	Target and CSHead are in a <i>conj</i> position via the same element, and stand in a <i>nsubj</i> position via the same word	(5.45)
b.	Target and CSHead are in a <i>conj</i> position via the same element, and stand in a <i>dobj</i> position via the same word	(5.46)
III.	Minor methods	(5.47), (5.48)

Table 5.17: Summary ISA1-Identifier

ISA-2 indicators are underlined:

(5.49) The Kuznets coal had a high energy content and low sulphur content, making it the *best* coal in the USSR.

(5.50) Russia remains the second *biggest* gold producer in the world.

[*Siberian natural resources*]

(5.51) Australia has many amateur companies, including the Adelaide Repertory Theatre, which claims to be the *oldest* in the country.

[*Amateur theatre*]

(5.52) Some prosthetic legs and feet allow for runners to adjust the length of their stride which could potentially improve run times and in time actually allow a runner with prosthetic legs to be the *fastest* in the world.

[*Cyborg*]

(5.53) He was the first writer to assume that the Caspian Sea was separated from other seas and he recognised northern Scythia as one of the *coldest* inhabited lands in the world.

[*History of cartography*]

- (5.54) Trier in Rhineland-Palatinate, whose history dates to the Roman Empire, is often claimed to be the *oldest* city in Germany.

[*History of Trier*]

- (5.55) He is regarded as one of the *best* coaches in Europe and in the Euroleague.

[*Ettore Messina*]

- (5.56) The *oldest* of these is called territory scoring and is used in Japan, Korea and most Western nations.

[*Go (board game)*]

- (5.57) As the world's *greatest* industrial power, and as one of the few countries physically unscathed by the war, the United States stood to gain enormously from opening the entire world to unfettered trade.

[*Origins of the Cold War*]

For each of these types, ISA2-Identifier has a set of syntactic rules based on the GR output of the C&C parser. For example, the output produced by the parser for (5.50) includes the following relations:

Tuple	GR
1	(dobj remains_1 producer_6)
2	(ncsubj remains_1 Russia_0 _)

To recognise ISA-2 superlatives of this type, the Identifier searches the GR output for tuples which satisfy the following requirements: 1.) the CSHead (*producer* in 5.50) stands in a dobj relation via a verb that occurs in the respective lexicon of ISA-2 indicators; and 2.) this verb is part of another tuple where it stands in an ncsubj relation with another word (which happens to be the target, *Russia*). A more complex search is carried out for cases like (5.54), where the relevant part of the GR output is as follows:

<b>Tuple</b>	<b>GR</b>
1	(xcomp - be_16 city_19)
2	(xcomp to_15 claimed_14 be_16)
3	(aux claimed_14 is_12)

Here, the Identifier first searches for a tuple where the CSHead stands in an `xcomp` relation via the verb “be” (Tuple 1). It then searches for a second tuple where “be” occurs in an `xcomp` relation via the complementiser “to”, and the main verb position is occupied by a lemma that occurs in the corresponding lexicon (Tuple 2). Finally, a third search is carried out to confirm that the ISA-2 indicator is used in passive voice (‘aux’ relation, Tuple 3). The other examples above involve similar strategies.

Table 5.18 shows a summary of the main patterns captured by ISA2-Identifier.

<b>Pattern</b>	<b>Short description</b>	<b>Example lexicon entries</b>	<b>Example</b>
I.	verb + T + CS	<i>make, consider</i>	(5.49)
II.	T + verb + CS	<i>become, constitute, remain</i>	(5.50)
III.	T + verb + <i>to be</i> + CS	<i>appear, believe, prove</i>	(5.51)
IV.	verb + T + <i>to be</i> + CS	<i>allow, cause</i>	(5.52)
V.	verb + T + <i>as</i> + CS	<i>consider, identify, recognise</i>	(5.53)
VI.	T (CS) + passive verb + <i>to be</i> + CS (T)	<i>claim, rumour, think</i>	(5.54)
VII.	T (CS) + passive verb + ( <i>as</i> ) + CS (T)	<i>consider, rate, regard</i>	(5.55)
VIII.	CS (T) + passive/non-finite verb + T (CS)	<i>call, vote</i>	(5.56)
IX.	<i>as</i> + CS	<i>as</i>	(5.57)

Table 5.18: Summary ISA2-Identifier

### 5.3.3.9 DEF

DEF-Identifier aims to recognise superlatives that are bound in a definite NP whose target is not explicitly mentioned in the same sentence (cf. Chapter 3.2.2.2). As DEF-Identifier is the last component in the classification pipeline, one could simply classify any instances that reach this level as DEF, since all other identifiers must have

failed. However, such an approach would result in a lower precision rate, as it would also capture instances that the preceding components in the pipeline falsely failed to identify. For this reason, a strategy for identifying DEF cases is described below, and any instances that are not captured are given a 0 (“null”) label for “unclassified”.

The procedure for identifying members of the DEF class is straightforward: The Identifier first checks whether `findHead` has identified a `CSHead`. If the result is positive, it checks the results of `findDeterminer`. If this test also succeeds, the instance is taken to occur in a definite NP and is labelled as DEF (5.58). Otherwise, it receives the label “0”.

(5.58) The<sub>{DET}</sub> *most* aggressive form<sub>{CSHead}</sub> of treatment involves surgical removal of the stone, via oral curette or a tonsillectomy to remove the tonsils.

[*Tonsillolith*]

Table 5.19 provides a summary of DEF-Identifier’s method.

Method	Short description	Example
I.	Check for presence of <code>CSHead</code> and <code>determiner</code>	(5.58)

Table 5.19: Summary DEF-Identifier

### 5.3.4 Results

SUP-Classifier is evaluated on the TextWiki test set, which contains 410 superlative sentences altogether. Seventeen of these sentences are excluded from evaluation as the C&C parser failed to parse them, leaving a total of 393 sentences for evaluation. The performance of SUP-Classifier is compared to a random baseline, which used a random generator to choose one of the nine class labels for each superlative in the test set. This achieved an average performance of 8.9%. Table 5.20 shows that the results of the random classifier for the individual classes range between 17.7% F-measure for the ISA-1 class (the second most frequent class) and 0% for FREE (the second least frequent class).

A second baseline system, which simply assigns the most frequent class label (“PROP”) to each superlative instance, achieves 45.3% F-measure for the PROP class, but 0% for

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F-Measure</b>
<b>IDIOM</b>	<b>3.8%</b> (2/53)	<b>18.2%</b> (2/11)	<b>6.3%</b>
<b>PP</b>	<b>3.8%</b> (2/53)	<b>9.1%</b> (2/22)	<b>5.4%</b>
<b>PROP</b>	<b>26.7%</b> (12/45)	<b>10.4%</b> (12/115)	<b>15.0%</b>
<b>ADV</b>	<b>7.1%</b> (3/42)	<b>9.7%</b> (3/31)	<b>8.2%</b>
<b>FREE</b>	<b>0%</b> (0/34)	<b>0%</b> (0/5)	<b>0%</b>
<b>INDEF</b>	<b>2.5%</b> (1/40)	<b>25.0%</b> (1/4)	<b>4.5%</b>
<b>ISA-1</b>	<b>25.5%</b> (12/47)	<b>13.5%</b> (12/89)	<b>17.7%</b>
<b>ISA-2</b>	<b>11.1%</b> (5/45)	<b>14.3%</b> (5/35)	<b>12.5%</b>
<b>DEF</b>	<b>17.6%</b> (6/34)	<b>7.4%</b> (6/81)	<b>10.4%</b>
<b>TOTAL</b>	<b>10.9%</b>	<b>12.0%</b>	<b>8.9%</b>

Table 5.20: Random baseline results

all other classes, and therefore has an overall F-measure of only 5.0%.

Table 5.21 shows the results of SUP-Classifier, which clearly outperforms both baselines with an overall F-measure of 77.0% (80.6% precision and 75.5% recall). There are no unclassified instances (Class “0”). In general, one can see that the four Identifiers whose methods rely mainly on lexical features (IDIOM, PP, PROP, ADV) perform better in terms of F-measure than the ones that rely on the output of the GR parser (FREE, INDEF, ISA-1, ISA-2, DEF). Quite strikingly, PP and PROP not only have the highest F-measure values of all (95.5% and 97.4%, respectively), but they also come first in terms of precision and recall, and can therefore be taken to be very reliable in performance. The classes FREE and INDEF show relatively poor results, with 60.0% and 40.0% F-measure. However, when interpreting these values one needs to bear in mind their relatively low frequency not only in the test set (five and four instances out of 393 in the test set), but also in the development set (14 and 8, respectively). This suggests that these sets were not sufficiently representative of the population as a whole, resulting in instances in the test set of a kind that had not been encountered during development.

The results of the remaining classes reflect that development focused on precision: The classes ISA-1, ISA-2, ADV and IDIOM all have very high precision values, ranging from 82.4% (ISA-1) to 100% (IDIOM). While ISA-1 also has a relatively high recall value (84.3%), the classes ISA-2, ADV and IDIOM have lower recall values (57.1%,

77.4% and 72.7%, respectively), most likely due to their dependence on various lexicons (see discussion below). Finally, the DEF class results reflect its position as the final component in the pipeline: They are characterised by high recall, but lower precision, its false positives coming from instances that the earlier components in the pipeline failed to identify.

<b>Class</b>	<b>Precision</b>	<b>Recall</b>	<b>F-Measure</b>
<b>IDIOM</b>	<b>100%</b> (8/8)	<b>72.7%</b> (8/11)	<b>84.2%</b>
<b>PP</b>	<b>95.5%</b> (21/22)	<b>95.5%</b> (21/22)	<b>95.5%</b>
<b>PROP</b>	<b>97.4%</b> (112/115)	<b>97.4%</b> (112/115)	<b>97.4%</b>
<b>ADV</b>	<b>92.3%</b> (24/26)	<b>77.4%</b> (24/31)	<b>84.2%</b>
<b>FREE</b>	<b>60.0%</b> (3/5)	<b>60.0%</b> (3/5)	<b>60.0%</b>
<b>INDEF</b>	<b>33.3%</b> (2/6)	<b>50.0%</b> (2/4)	<b>40.0%</b>
<b>ISA-1</b>	<b>82.4%</b> (75/91)	<b>84.3%</b> (75/89)	<b>83.3%</b>
<b>ISA-2</b>	<b>95.2%</b> (20/21)	<b>57.1%</b> (20/35)	<b>71.4%</b>
<b>DEF</b>	<b>69.7%</b> (69/99)	<b>85.2%</b> (69/81)	<b>76.7%</b>
<b>Average</b>	<b>80.6%</b>	<b>75.5%</b>	<b>77.0%</b>

Table 5.21: SUP-Classifier results

### 5.3.5 Discussion

In order to assess the performance of SUP-Classifier, a thorough analysis of the classification results has been carried out, resulting in the confusion matrix in Table 5.22. The rows of the table correspond to the gold standard classes, while the columns correspond to the predicted classes. For example, the first row labelled “IDIOM” shows that the Classifier predicted 8 instances correctly as IDIOM, but falsely labelled one instance as FREE, another one as ISA-1, and one as DEF. The sum of the row values corresponds to the total number of instances labelled as IDIOM in the gold standard (11).

For a fair evaluation of the individual components, the results presented in Table 5.21 should be interpreted in the context of the cascade used in processing the superlative sentences, since each subsequent component of the cascade sees fewer and fewer examples. Thus, for each component, its set of False Negatives may include items it never gets to see, as they are False Positive of some component earlier in the cascade.

Row	Class	IDIOM	PP	PROP	ADV	FREE	INDEF	ISA-1	ISA-2	DEF
1	<b>IDIOM</b>	<b>8</b>	0	0	0	1	0	1	0	1
2	<b>PP</b>	0	<b>21</b>	1	0	0	0	0	0	0
3	<b>PROP</b>	0	0	<b>112</b>	0	0	2	0	0	1
4	<b>ADV</b>	0	0	0	<b>24</b>	1	2	3	0	1
5	<b>FREE</b>	0	0	0	2	<b>3</b>	0	0	0	0
6	<b>INDEF</b>	0	0	2	0	0	<b>2</b>	0	0	0
7	<b>ISA-1</b>	0	0	0	0	0	0	<b>75</b>	0	14
8	<b>ISA-2</b>	0	0	0	0	0	0	2	<b>20</b>	13
9	<b>DEF</b>	0	1	0	0	0	0	10	1	<b>69</b>

Table 5.22: Confusion matrix

Therefore, to give a picture of each component separately, one needs to consider what cases it would be presented with if all earlier components in the cascade had performed perfectly, and whether it would have handled these cases correctly or not.

The confusion matrix can be used to visualise this problem. The diagonal cells highlighted in bold count the correctly identified instances. All cells to the left of this line (“lower left triangle”) represent cases which were classified as False Positives by another component earlier in the cascade. The cells to the right of the diagonal line (“upper right triangle”) contain all instances whose respective Identifier failed, and which were subsequently classified by another component. A comparison of the numbers of errors within the two triangles shows that only 18 out of 59 incorrectly classified instances (30.5%) were classified as False Positive of some component earlier in the cascade, while the remaining 41 errors occurred due to failure of their respective Identifier (69.5%). This reflects the fact that the methods described in Section 5.3.3 were developed to achieve high precision (rather than recall).

Table 5.23 shows precision and recall values for the individual Identifiers given the scenario that all preceding Identifiers had perfect precision and recall.<sup>9</sup> The classes that were able to improve their recall values are FREE, INDEF, and DEF, as all instances that were False Positives of some component earlier in the cascade were correctly recognised. This does not apply to ISA-2: It did not succeed in identifying the two instances that had been recognised as ISA-1 under normal conditions, and its recall value therefore remains the same.

<sup>9</sup>New results have been underlined.

With 100% precision and recall assumed for earlier components, the evaluation shows especially good results for the FREE, INDEF and DEF classes. While this may not be surprising for DEF since it is the last component in the cascade, the results indicate that the FREE and INDEF Identifiers' methods perform well on their own, and that the problem lies in the methods of the preceding components.

Class	Precision	Recall	F-Measure
<b>IDIOM</b>	<b>100%</b> (8/8)	<b>72.7%</b> (8/11)	<b>84.2%</b>
<b>PP</b>	<b>95.5%</b> (21/22)	<b>95.5%</b> (21/22)	<b>95.5%</b>
<b>PROP</b>	<b>98.2%</b> (112/114)	<b>97.4%</b> (112/115)	<b>97.8%</b>
<b>ADV</b>	<b>92.3%</b> (24/26)	<b>77.4%</b> (24/31)	<b>84.2%</b>
<b>FREE</b>	<b>100%</b> (5/5)	<b>100%</b> (5/5)	<b>100%</b>
<b>INDEF</b>	<b>100%</b> (4/4)	<b>100%</b> (4/4)	<b>100%</b>
<b>ISA-1</b>	<b>86.2%</b> (75/87)	<b>84.3%</b> (75/89)	<b>85.2%</b>
<b>ISA-2</b>	<b>95.2%</b> (20/21)	<b>57.1%</b> (20/35)	<b>71.4%</b>
<b>DEF</b>	<b>100%</b> (81/81)	<b>100%</b> (81/81)	<b>100%</b>

Table 5.23: SUP-Classifier results

The following paragraphs will use the confusion matrix to investigate the performance of the individual components of SUP-Classifier in more detail.

**IDIOM-Identifier**, described in Section 5.3.3.1, makes use of three main methods, the CapsTest method, the IdiomLex lexicon, and a check for particular syntactic patterns. An investigation of the results shows that of the eight instances that were correctly predicted, half were recognised by CapsTest, and the other half by IdiomLex (cf. Table 5.24, row “IDIOM”). With a precision value of 100%, one can conclude that these methods can identify IDIOM instances very reliably. Recall, however, is lower at 72.7%, and Row 1 in the confusion matrix shows that three IDIOM instances were falsely labelled as ISA-1 (5.59), DEF (5.60), and FREE (5.61)<sup>10</sup>:

- (5.59) According to FIFA, the “very *earliest* form of the game for which there is scientific evidence was an exercise of precisely this skilful technique dating back to the 2nd and 3rd centuries BC in China (the game of cuju).

[Association football]

<sup>10</sup>Superlative forms have been italicised in the examples.

- (5.60) For the *most* part, they are bilaterally symmetric, and often have a specialized head with feeding and sensory organs.

[*Animal*]

- (5.61) The Portuguese government were not in the *least* bit reluctant to part with Tangier.

[*Tangier Garrison*]

Example (5.59), which has been classified as IDIOM due to the modifier “very”, should have been identified via the syntactic rules. So far, these do not consider superlatives preceded by “very”, because they are also frequently part of PP constructions (cf. 5.3.3.2). An updated version of SUP-Classifier should include an additional syntactic rule for cases like (5.59).

The words “most” and “least” in examples (5.60) and (5.61) are both part of fixed idiomatic expressions, and are missing from IdiomLex. The updated version of SUP-Classifier will include the phrases “for the most part” and “the least bit” in its Idiom Lexicon.

Row 2 in Table 5.22 shows that out of 22 PP instances in the data set, **PP-Identifier** failed to recognise only one, which was classified as PROP (5.62):

- (5.62) This is further backed up by the fact that the Orungu seemed to have been heavily influenced by the Kingdom of Loango or at very *least* its BaVili traders.

[*Kingdom of Orungu*]

Case (5.62) was not recognised by PP-Identifier’s Method 2 due to a missing definite article (cf. Table 5.8). SUP-Classifier subsequently labelled the instance as PROP, because the supertag of the superlative is “N”, and the tag of the next token is not RB (listed among the “Rules of inclusion”, cf. Section 5.3.3.3). Table 5.24 shows that most other PP instances were identified correctly via the simple rule “at” + “least” (90.5% altogether).

**PROP-Identifier** recognised 112 out of 115 instances correctly (Row 3 in Ta-

ble 5.22). Its first method, “Rules of exclusion”, was applied 97 times, successfully identifying and excluding clear non-PROP cases. Table 5.24 also shows that Methods II and III accounted for 94 (83.9%) and 18 (16.1%) out of all identified instances. PROP-Identifier only labelled three instances incorrectly, two as INDEF (5.63 and 5.64) and one as DEF (5.65):

- (5.63) Given the premise that some, if not *most*, adolescents are going to have sex, a harm-reductionist approach supports a sexual education which emphasizes the use of protective devices like condoms and dental dams to protect against unwanted pregnancy and the transmission of STDs.

[*Harm reduction*]

- (5.64) Yet, fortification is used regularly as a marketing strategy, so that now *most* processed foods are fortified in some way.

[*Nutrition*]

- (5.65) The development of *most* the famous gold-field at Muruntau began in the early 1960s.

[*Kyzyl Kum*]

The misclassification of (5.65) is clearly due to an error in the input sentence. The intended version is likely to be “the most famous gold-field” or “most of the famous gold-fields”. Strictly seen, this instance should have been excluded from evaluation, as its gold standard classification of PROP is questionable due to the ungrammaticality of the sentence. Example (5.63) was not recognised by PROP-Identifier because of the appositive position of “most”, while classification of (5.64) as PROP failed due to the length of the adjective following “most”: Method III of PROP-Identifier (“Strategies for ‘most’/‘least’ + JJ/VBD/VBN”) set the threshold to < 6 characters (cf. 5.3.3.3). As the word “processed” has length 9, PROP-Identifier fails, and the instance is subsequently classified as INDEF.

**ADV-Identifier** correctly identified 24 out of 31 ADV instances (Row 4 in Table 5.22). While most of them were recognised by the set of inclusion tests (83.3%), ADV-Identifier’s other methods also accounted for a number of instances (see row

labelled “ADV” in Table 5.24). The seven incorrect predictions are spread across the FREE, INDEF, ISA-1, and DEF classes. Four out of seven errors were caused by the pattern “most” + deverbal adjective, as in (5.66) and (5.67), which were classified as DEF and ISA-1, respectively.

- (5.66) The *most* cited<sub>{VBD}</sub> reasons for introducing massage was patient demand and perceived clinical effectiveness.

[*Massage*]

- (5.67) For over two decades, the Charnley Low Friction Arthroplasty design was the *most* used<sub>{JJ}</sub> system in the world, far surpassing the other available options (like McKee and Ring).

[*Hip replacement*]

The inclusion test “most” + tag VBN (Method I), which is responsible for identifying cases like the above (cf. 5.3.3.4), failed in (5.66) due to a tagging error: The word “cited” is incorrectly tagged as VBD (verb, past tense). In (5.67), on the other hand, the word “used” is treated by the tagger as a full adjective JJ, while the gold standard annotation considers it a deverbal adjective and classifies the instance as ADV (cf. discussion in Chapter 3.2.3).

The remaining three errors result from another tagging error, where *best* is tagged as JJS and *expresses* as NNS (instead of VBZ) (5.68), and an unhyphenated occurrence of *longest* modifying a deverbal adjective (*established*) in (5.69), which is not yet included in the list of lemma-based searches described in 5.3.3.4 (Method III). Finally, example (5.70) describes a syntactic pattern that is not listed among ADV-Identifier’s Method IV. It was excluded deliberately during development as it produced too many false positives in the development sets.<sup>11</sup>

- (5.68) Legend of Mana’s score was composed by Yoko Shimomura, and of all her compositions, she considers it the one that *best*<sub>{JJS}</sub> *expresses*<sub>{NNS}</sub> herself.

[*Mana (series)*]

<sup>11</sup>The relevant GR tuple is (xcomp \_ fare\_20 best\_21).

- (5.69) Dollond & Aitchison are the United Kingdom's *longest* established opticians, having been established in 1750 by Peter Dollond in Vine Street, London, and been joined by his father John Dollond 2 years later.

[*Dollond & Aitchison*]

- (5.70) John Forester, a cycling transportation engineer, has written that the principle of vehicular cycling is: "Cyclists fare *best* when they act and are treated as drivers of vehicles".

[*Vehicular cycling*]

Of the five instances classified as FREE in the gold standard, three were identified correctly by **FREE-Identifier**, two by Method I and one by Method II (Row 1 in Table 5.25). The other two FREE instances were incorrectly classified as ADV in the previous step of the classification pipeline (Row 6 in the confusion matrix). The first of these, example (5.71), is actually a borderline case between ADV and FREE, as *associated* is a deverbal adjective. It therefore could also have been classified as ADV in the gold standard. The second incorrectly classified instance is shown in (5.72), which again could have been labelled as ADV in the gold standard, depending on whether *well known* is considered a deverbal adjective or a lexicalised compound adjective.

- (5.71) Work *most* associated with psychologist Albert Bandura, who initiated and studied social learning theory, showed that children could learn by social observation, without any change in overt behavior, and so must be accounted for by internal representations.

[*Psychology*]

- (5.72) In The Netherlands the Dada movement centered mainly around Theo van Doesburg, *most* well known for establishing the De Stijl movement and magazine of the same name.

[*Dada*]

As Table 5.22 shows (Row 6), there are only four INDEF instances altogether in the test set, of which two were correctly identified by **INDEF-Identifier**. The other two instances were classified as PROP earlier on in the pipeline:

- (5.73) In general they conclude that animal complete proteins that contain all the essential amino acids such as milk, eggs, and meat, and the complete vegetable protein soy are of *most* value to the body.

[*Protein in nutrition*]

- (5.74) The wealthy, who may already eat enough nutrients, may be consuming more than they need when they purchase fortified products, while those in *most* need of the added nutrients may not be able to afford to buy enough fortified foods.

[*Nutrition*]

At a closer look, the surface structures of (5.73) and (5.74) are ambiguous between a superlative reading and proportional quantification. The gold standard annotation as INDEF indicates that in both cases “most” should be taken to express a superlative comparison. For example, in (5.73) the comparison is between different proteins and the strength of their value to the body. Cases like these cannot be recognised by INDEF-Identifier, as they match one of the PROP inclusion rules (“most”/“least” + NNS or NN), which is applied at an earlier stage in the classification pipeline.

**ISA1-Identifier** recognised 75 out of 89 ISA-1 instances in the test set. Table 5.25 shows that the great majority of these (64/75 or 85.3% altogether) were recognised by Methods Ia (41.3%) and Ib (44.0%), which address ISA-1 instances whose relation between target and comparison set is expressed via the verb “to be”. The remaining 11 instances were identified via the `nsubj` apposition rule (IIa, 2.7%) and the `dobj` apposition rule (IIb, 9.3%), and via the minor methods (cf. Section 5.3.3.7).

According to the confusion matrix in Table 5.22 (Row 7), all 14 ISA-1 instances that were not recognised by the Identifier were subsequently classified as DEF. A closer investigation of them (a complete list can be found in Appendix C.1 along with a short note explaining the error) shows that the majority (12/14) involve appositions which were either not recognised by the parser (5.75), or not recognised by ISA1-Identifier

due to failure of one of its methods (5.76), or because no strategy is yet available for recognising particular types of appositions (5.77).

- (5.75) The *largest* animal phylum belongs here, the Arthropoda, including insects, spiders, crabs, and their kin.

[*Animal*]

- (5.76) The *highest* point in the archipelago, Conachair ('the beacon') at 430 metres (1,411 ft), is on Hirta, immediately north of the village.

[*St Kilda, Scotland*]

- (5.77) When Baldwin IX, Count of Flanders and Hainault, left on the Fourth Crusade in 1202, he left his western domains under his *eldest* daughter Joanna.

[*War of the Succession of Flanders and Hainault*]

In the GR output for (5.75), there are no tuples that describe a conjunction relation (`conj`) involving the comma with index 6. This is most likely due to the unusual position of the apposition “the Arthropoda” in the sentence: Rather than following the `CSHead` *phylum* directly, it is attached at the end of the clause, after the word “here”. In (5.76), on the other hand, the parser identified the `conj` relations between the `CSHead` *point* and the noun *Conachair*, namely (`conj ,_6 Conachair_7`) and (`conj ,_6 point_2`), but `ISA1-Identifier` does not recognise their shared verb “is”, and therefore fails. Finally, (5.77) is an example of a special kind of apposition, where the appositive *Joanna* is not set off by commas, but follows the `CSHead` directly (a so-called “restrictive” apposition). All strategies for this type that were experimented with during development (e.g. involving Named Entity annotations) lowered precision considerably. Due to the relatively low frequency of such instances, I therefore decided to exclude this type from consideration.

Column 8 in the confusion matrix shows that **ISA-2 Identifier** has a high precision rate: Only one out of 21 assigned ISA-2 labels was incorrect (should have been classified as DEF). This result suggests that the output of the pattern-based approach described in Section 5.3.3.8 is very reliable. Table 5.25 shows that a variety of patterns occurred in the test data, with cases like (5.78) and (5.79) being among the most

common.

- (5.78) El Bulli has 3 Michelin stars and is regarded as one of the *best* restaurants in the Western world.

[*Ferran Adri*]

- (5.79) The Discovery Channel’s Young Scientist Challenge began in 1999 as Science Service’s *newest* innovation.

[*Science Service*]

ISA-2 recall is lower, with 15 incorrectly predicted instances. Two ISA-2 superlatives were wrongly classified as ISA-1 in the preceding step of the pipeline, while the other 13 were not recognised by ISA-2 Identifier and received DEF labels in the subsequent step (cf. confusion matrix, row 8). Strikingly, none of these errors were due to an unknown pattern, which shows that the list of nine types of ISA-2 constructions can be considered sufficient to describe the class. The errors can be broadly divided up into three groups: The first one includes cases which match one of the nine patterns, but whose ISA-2 indicator is missing from the appropriate lexicon. A list of all cases is provided in Appendix C.2. Altogether, the following five types are missing from their respective lexicon:

- *call*, as in “*call* + T + CS” (Lexicon I)
- *know*, as in “CS + *is known* + (*as*) + T” (Lexicon VII)
- *show*, as in “*show* + T + *to be* + CS” (Lexicon IV)
- *seem*, as in “T *seems* CS” (Lexicon II)
- *consider*, as in “T *considered* CS” (Lexicon VIII)

The second group of ISA-2 instances that were not recognised by the Identifier contains cases like (5.80), where the CS is preceded by “among” (3 instances altogether), and (5.81), which contains an “of which” construction. Both of these are dealt with by ISA1-Identifier (Method III), but are not considered by ISA-2 Identifier.

- (5.80) Venezuela is considered to be among 17 of the *most* megadiverse countries in the world.

[Venezuela]

- (5.81) The chain consists of two large islands (Aorangi and Tawhiti Rahi) with a group of smaller islets between the two, the *largest* of which is called Motu Kapiti.

[Poor Knights Islands]

Finally, there is one ISA-2 superlative that was not recognised because of an error produced by the coordination detection module `findCoord` (6.30).

- (5.82) The whole north face of Conachair is a vertical cliff up to 427 metres (1,400 ft) high, falling sheer into the sea and constituting the *highest* sea cliff in the UK.

[St Kilda, Scotland]

The last component in the pipeline, **DEF-Identifier**, is the only one that is characterised by a precision rate that is clearly lower than its recall (69.7% vs. 85.2%). Considering the fact that `DEF-Identifier` only uses one method, it works very well in retrieving all remaining DEF instances that reach this point in the pipeline. It is also striking that despite using this method, there are no superlative instances that remain unclassified (Label “0”). The confusion matrix shows that 12 instances were not identified as DEF. Of these, one was classified as PP (5.83), as it matches the pattern “in” + “the” + B-SUP. Another DEF superlative was classified as ISA-2 (5.84), and represents the only case where an ISA-2 label was assigned mistakenly (cf. column labelled “ISA-2” in Table 5.22.)

- (5.83) In this case, we have a visual cue with the dolls for Russia (as the plot involves Soviet espionage), as well as with the final doll for the unknown mole, a spy who’s buried in the *deepest*.

[Matryoshka doll]

- (5.84) Despite their reputation as the *most* prominent [*sic*] slave traders in the region, some visitors to the kingdom left favorable reviews of the region and

its people.

[*Kingdom of Orungu*]

The other 10 instances were falsely classified as ISA-1, and are listed in Appendix C.3. They can broadly be divided up into two categories: Clear errors with no target entity, as for example (5.85), and borderline cases, where the decision of `ISA1-Identifier` to classify the instance as ISA-1 could to some degree be justified. An example of this is (5.86), where the string “a straight channel” could be considered to represent a target.

- (5.85) The *most* serious symptoms of altitude sickness are due to edema (fluid accumulation in the tissues of the body).

[*Altitude sickness*]

- (5.86) The *shortest* distance; that is, a straight channel, results in the highest energy per unit of length, disrupting the banks more, creating more sediment and aggrading the stream.

[*Meander*]

### 5.3.6 Conclusion

The results in Table 5.21 and the error analysis in the last section suggest that `SUP-Classifier` is a tool that can be used to distinguish between different superlative classes with very good results. The rule-based approach combined with a cascading architecture has the advantage that each class can be addressed separately in terms of their distinguishing features. The analysis in the previous section has shown that there is still room for improvement. For example, rules may be added for phenomena that did not occur in the development data, such as adding the “Minor Methods” of `ISA1-Identifier` to the set of methods of `ISA2-Identifier`. Furthermore, errors that result from unanticipated tagging (or parsing) errors could be taken into account by adding further rules to the individual Identifiers. As a large proportion of the errors occurred in connection with Identifiers that rely on lexicon-based searches, another

way of improving the system is by adding missing items to the respective lexicons to improve coverage. This particularly applies to the IDIOM, ADV, and ISA-2 Identifiers. Other errors, on the other hand, cannot be resolved with the means at hand, as they involve ambiguities that could only be resolved with a sophisticated semantic analyser (if at all). A case in point is for example the ambiguity between the readings of *most* as proportional quantifier and as superlative (5.73 and 5.74). As with most natural language processing tasks, some of these cases may not be resolvable at all, thus placing an upper limit to the performance of the classifier.

Class	Method	Short description	Test set matches
IDIOM	I.	CapsTest: capitalisation of B-SUP?	4/8 (50%)
	II.	IdiomLex: listed in Idiom Lexicon?	4/8 (50%)
	III.	Lexical/syntactic pattern?	0/8 (0%)
PP	I.	["at" + "least"]	19/21 (90.5%)
	II.	["at" + B-SUP] $\wedge$ [I-SUP + $\neg$ {NN, NNS, NNP, "of"}]	1/21 (4.8%)
	III.	["at"/"in" + "the" + ("very") + B-SUP] $\wedge$ [I-SUP + $\neg$ {NN, NNS, NNP, JJ, "of"}]	1/21 (4.8%)
PROP	I.	Rules of exclusion	excluded: 97
	II.	Rules of inclusion	94/112 (83.9%)
	III.	Strategies for "most"/"least" + JJ or VBD/VBN	18/112 (16.1%)
ADV	I.	Inclusion tests	20/24 (83.3%)
	II.	Compound search: I-SUP + hyphen	1/24 (4.2%)
	III.	Lemma-based searches	1/24 (4.2%)
	IV.	Syntactic patterns	2/24 (8.3%)

Table 5.24: Performance of Identifiers in SUP-Classifier Part I

Class	Method	Short description	Test set matches
FREE	I.	findHead fails	2/3 (66.7%)
	II.	findHead has identified a fused head; <i>findDeterminer</i> has found no determiner	1/3 (33.3%)
INDEF	I.	findHead succeeds, findDeterminer fails	2/2 (100%)
	II.	findHead and findDeterminer succeed, determiner is indefinite (“a” or “an”)	0/2 (0%)
ISA-1	I a.	CSHead in nsubj position via “to be”	31/75 (41.3%)
	b.	CSHead in xcomp position via “to be”	33/75 (44.0%)
	II a.	Target and CSHead in conj position via same element, and stand in nsubj position via same word	2/75 (2.7%)
	b.	Target and CSHead in conj position via same element, and stand in dobj position via same word	7/75 (9.3%)
	III.	Minor methods	2/75 (2.7%)
ISA-2	I.	verb + T + CS	2/20 (10.0%)
	II.	T + verb + CS	4/20 (20.0%)
	III.	T + verb + <i>to be</i> + CS	2/20 (10.0%)
	IV.	verb + T + <i>to be</i> + CS	0
	V.	verb + T + <i>as</i> + CS	0
	VI.	T (CS) + passive verb + <i>to be</i> + CS (T)	0
	VII.	T (CS) + passive verb + ( <i>as</i> ) + CS (T)	6/20 (30.0%)
	VIII.	CS (T) + passive/non-finite verb + T (CS)	0
	IX.	“as” CS	6/20 (30.0%)
DEF	I.	Check for presence of CSHead and determiner	69 (100%)

Table 5.25: Performance of Identifiers in SUP-Classifier Part II

# Chapter 6

## Superlative Relation Extraction

### 6.1 Overview

This chapter describes an experiment that deals with the automatic identification of the target (T) and comparison set (CS) strings of superlatives classified as ISA-1 in the previous step (6.2). The resulting program, called “T/CS-Identifier”, can be combined with the other two components to create a “Superlative Relation Extraction” (SRE) pipeline. All components are developed and tested on the `TextWiki` data sets described in Chapter 4.4.

### 6.2 Experiment 3: Identifying the target and comparison set of ISA-1 superlatives (T/CS-Identifier)

#### 6.2.1 Overview

The third experiment deals with the automatic identification of the target (T) and comparison set (CS) strings of ISA-1 superlatives. An initial investigation of the output produced by the C&C parser showed that an approach based on GR and supertags (CCG categories) suffers from a number of shortcomings. Bos and Nissim (2006), whose study of superlatives addressed the problem of comparison set detection (cf. Chapter 2.2.2), describe similar problems. Their system is based on a semantic representation of the superlative construction in terms of Discourse Representation Structure

(DRS). This approach relies on the output of the C&C parser, whose output (according to Bos and Nissim) is not always sufficient to construct a meaningful semantic representation. They note that it is unable to handle NP post-modification of the superlative on the one hand, and possessive NPs preceding the superlative construction on the other, resulting in wrong semantic analyses for both of these issues. The first problem is caused by an attachment problem, illustrated by Bos and Nissim with the example “*the largest toxicology lab in New England*”. Here, the parser attaches the modifier to the NP node rather than N. The PP *in New England* is assigned the CCG category NP\NP instead of N\N, which implies that the comparison set consists of *toxicology labs*, rather than *toxicology labs in New England*. The second problem is illustrated with the example “*Jaguar’s largest shareholder*”, for which the parser outputs a derivation where “largest” is first combined with “shareholder”, and then with the possessive construction (rather than the other way round). This also produces a wrong semantic interpretation. The latter problem is analysed in detail in a recent paper by Bos (2009).

Bos and Nissim address these issues by using post-processing rules to alter the CCG derivation output by the parser, resulting in four different versions of their DLA (Deep Linguistic Analysis) system. The first one uses the unmodified DRS output of their system (DLA 1). The second and third versions build on DLA 1 by adding post-processing rules to the CCG derivations, to deal with NP post-modification of the superlative on the one hand (DLA 2), and with possessives preceding the superlative on the other (DLA 3). The fourth version, which is shown to perform best, combines both post-processing rules into one system (DLA 4).

As discussed in Chapter 2.2.2, Bos and Nissim’s study aims to identify the *span* of the comparison set without further analysis of the constituents of the resulting string. This means that no information will be available about what sort of entities the comparison set actually contains, and what sort of restrictions apply to it. For example, in “*the largest toxicology lab in New England*”, the set consists of toxicology labs and is restricted in location: Only the ones *in New England* are considered. More crucially, their system does not distinguish between different types of superlative comparison. This means that no information is available about whether the target of comparison is mentioned in the same sentence, and if so, where it is located.

The goal of the present study is to extend and improve on Bos and Nissim’s approach in three ways. First, with the aid of the superlative classification system described in Section 5.3, it is possible to identify superlatives whose target entity is mentioned in

the same sentence (ISA-1), making it possible to extract a complete superlative relation in terms of target (T), superlative relation, and comparison set (CS). Secondly, a new system is created for detecting the target and comparison set spans of ISA-1 superlatives (T/CS-Identifier), which also takes into account their internal structure. This approach, which can be refined in a later step, is particularly suitable for an analysis of the semantic roles of the individual CS constituents, and will allow for an interpretation of the heads and restrictions of ISA-1 superlatives. Thirdly, T/CS-Identifier also incorporates an analysis of the (extended) determinative phrase of the CS (cf. Chapter 4.2.2). Bos and Nissim only take possessive NPs into account, which usually restrict the comparison set (such as “*Finland’s largest company*”). However, the determinative phrase is also responsible for indicating the target’s position and uniqueness compared to the other members of the comparison set with respect to the property expressed by the superlative. For example, the determiner “*the (largest)*” implies that the target has indeed the highest value of the property expressed by the superlative, while “*the second (largest)*” indicates that there is at least one other member of the CS with a higher value. When comparison sets are preceded by the “extended” determinative phrases “*one of the (largest)*” or “*among the (largest)*”, the exact position of the target remains unspecified, while “*the two (largest)*” or “*the other (largest)*” imply non-uniqueness of the target(s).

## 6.2.2 Method

T/CS-Identifier is implemented in Python and added to the SUP-Classifier pipeline, as it makes use of its output for the ISA-1 class. In particular, it requires information about the superlative start and end index (B-SUP and I-SUP), the results of `findHead` (i.e. the CSHead index) and `findDeterminer` (i.e. the Det index), as well as information about the location of CSHead in the sentence: Subject position (`ncsubj`), complement position (`xcomp`), subject apposition (`subjapp`), or object apposition (`objapp`). In addition, T/CS-Identifier has access to `TagList` and `GRList` (cf. Section 5.3.3).

T/CS-Identifier consists of two parts: The first one tries to identify the span of the comparison set of a given ISA-1 superlative (CS-Identifier), and the second one attempts to locate and identify the span of the target (T-Identifier). Their goal is to identify all relevant constituents of the T and CS phrases (as defined in Chapter 4.2.2) by using a fine-grained set of rules, which are mainly based on information gathered

from `TagList`. The purpose of this syntactic analysis of CS constituents is not only the identification of the T and CS spans, but it also offers itself as basis for a second level of analysis in terms of their semantic roles, which would then allow for an extraction of CS restrictions.

The present task assumes that both target and comparison set comprise a single span. As discussed in Chapter 4.2.2, this assumption can be violated through preposing (as “in the USA” in (6.1)). However, for this investigation, CS span is assumed to comprise of a single string that is attached to the `CSHead`, and preposed elements or external modifiers are not taken into account.

(6.1) **In the USA** {licensure} is [the] *highest* [level of regulation] and this restricts anyone without a license from practicing massage therapy or by calling themselves that protected title.

[*Massage*]

### 6.2.3 Identifying the comparison set span (CS-Identifier)

The CS span (`CSspan`) is defined as consisting of a determinative phrase (`CSDet`) and the main CS phrase (`CSMain`) (cf. Chapter 4.2.2). `CS-Identifier` consists of two separate components that attempt to identify these strings, which are described in the following two sections.

#### 6.2.3.1 Identifying the determinative phrase (CSDet)

`T/CS-Identifier` first attempts to identify the determinative phrase, and uses a purely pattern-based approach. Starting from B-SUP position and moving to the left, `CS-Identifier` checks each token against the pattern displayed in Table 6.1. Six POS tags are taken into account: DT (determiner), POS (genitive marker), PRP\$ (possessive pronoun), CD (cardinal numeral), JJ (adjective or ordinal numeral), and RB (adverb). In addition, the lexical items *among* and *some* are also considered.

Comparing this pattern against the structure of the determinative phrase shown in Chapter 4.2.2 (Table 4.2), one can see that the central determiner slot (DET) is realised by the POS tags DT, POS or PRP\$ (as POS is an indicator of possessive 's,

it needs at least one preceding token chunked I-NP). While the central determiner is obligatory, the remaining components are optional. The central determiner may be preceded by the patterns “CD + of”, “some of”, and “among”, which can also occur before the former two (cf. Chapter 4.2.2). It can be followed either by a token tagged JJ (e.g. *other*, *next*, *second*) or CD (e.g. *two*, *three*, ...), or a combination of the two (e.g. *the other two largest buildings*). As JJ can also denote ordinal numerals, tokens like “second” and “third”, which should be interpreted as part of the superlative form, are excluded from the `CSDet` span. This is achieved by checking if the token in question equals “second” or “third”, or ends in “th”.

CSDet														
(among)	+	[(CD)	+	(of)]	+	[(I-NP)+	+	<b>DT</b>	+	(JJ)	+	(CD)	+	B-SUP
		[(some)						<b>POS]</b>		(CD)		(JJ)		
								<b>PRP\$</b>						

Table 6.1: CSDet pattern

The pattern can identify determiners like the following:

(6.2) DotA is also slated to appear in ESWC 2008, [one<sub>{CD}</sub> of<sub>{of}</sub> the<sub>{DT}</sub>] *most*<sub>{B-SUP}</sub> *prestigious* gaming events in electronic sports.

[*Defense of the Ancients*]

(6.3) From 318 onwards, Trier was the seat of the Gallic prefecture (the Praefectus Praetorio Galliarium), [one<sub>{CD}</sub> of<sub>{of}</sub> the<sub>{DT}</sub> two<sub>{CD}</sub>] *highest*<sub>{B-SUP}</sub> authorities in the Western Roman Empire, which governed the western Roman provinces from Morocco to Britain.

[*History of Trier*]

(6.4) General Grievous, Lobot, and Luke Skywalker are [the<sub>{DT}</sub> three<sub>{CD}</sub> other<sub>{JJ}</sub>] *most*<sub>{B-SUP}</sub> *prominent* cyborgs in the Star Wars universe.

[*Cyborg*]

Finally, `CS-Identifier` employs a back-off strategy for cases where the determinative phrase is not immediately followed by the superlative NP, as for example “most

popular” in (6.5). Starting with the token preceding B-SUP, the Identifier moves to the left, and for each token checks whether it matches DT, POS, or PRP\$. If it does, the Identifier adjusts the central determiner index, and proceeds to check for any of the preceding patterns shown in Table 6.1.

- (6.5) It is [one<sub>{CD}</sub> of<sub>{of}</sub> Bavaria<sub>{I-NP}</sub> ’s<sub>{POS}</sub>] oldest and *most*<sub>{B-SUP}</sub> popular festivals and it has increased to one of the biggest events in Germany.

[Gäubodenvolksfest]

### 6.2.3.2 Identifying the main CS span (CSMain)

The second part of CS-Identifier deals with the main CS span and consists of a variety of rules centered around CSHead, which was determined by the module findHead and is assumed to be the superlative NP head. The goal is to determine the span of the comparison set by identifying potential premodifiers (“PreMod”) and postmodifiers (“PostMod”) of CSHead. The main CS span CSMain is then taken to consist of the string [PreMod + CSHead + PostMod].

Identifying potential **premodifiers** is fairly straightforward: CS-Identifier simply includes all tokens that occur between I-SUP and CSHead (as in 6.6), unless it encounters the token “and”, or punctuation marks such as commas, brackets, or dashes (6.7). If there are no tokens between I-SUP and CSHead, the main CS span is taken to start with the token labelled CSHead (6.8). For fused head cases, where CSHead coincides with the token in I-SUP position, the main CS span starts with the token following CSHead.

- (6.6) Reservoir Park is the *largest*<sub>{I-SUP}</sub> [municipal public]<sub>{PreMod}</sub> park<sub>{CSHead}</sub> in Harrisburg, Pennsylvania, and occupies approximately 85-acres in the Allison Hill neighborhood of the city.

[Reservoir Park (Harrisburg)]

- (6.7) The *most common*<sub>{I-SUP}</sub> (and still ongoing) [problem<sub>{CSHead}</sub>], though, is that people sift through the Free Box without any regard for keeping it neat and organized.

[Bolinás Free Box]

(6.8) The *lowest*<sub>{I-SUP}</sub> [class<sub>{CSHead}</sub> of people] are the da'covale, “those who are property.”

[Seanchan]

Finding all relevant **postmodifiers**, on the other hand, is a complex task. In Chapter 4.3.2 I noted that slightly over two thirds of the ISA-1 superlatives in the pilot study (60/89) have a comparison set with at least one postmodifier. Unlike in the case of identifying premodifiers, whose span is limited by the superlative form to the left and CSHead to the right, there is no delimiter on the right hand side of CSHead.

Chapter 4.2.2 provided the following list of types of postmodifiers:

- I. PP: prepositional phrase
- II. ADVP: adverbial phrase
- III. ADJP: adjectival phrase
- IV. CL-NONF: non-finite clause
- V. CL-REL: restrictive relative clause (non-restrictive ones are excluded)
- VI. NP: noun phrase
- VII. CL-SUB: subordinate clause

Given a token following CSHead, the current approach assumes that the following three main questions need to be addressed to identify it as postmodifier of CSHead:

- Question 1: Does the token introduce a postmodifying phrase or not?
- Question 2: If it does, is the phrase restrictive or not?
- Question 3: If it is restrictive, what is its exact span?

To solve **Question 1**, I devised a fine-grained set of rules to match the types of postmodifiers shown above as closely as possible (“PostMod” rules). An overview of these is shown in Table 6.11, and an individual description of each rule is given in Appendix D.1. The Identifier starts out with the first token to the right of CSHead, and checks if it matches the “trigger” token of any of the PostMod rules shown in Table 6.11 at the end of this chapter. If it does, the token is assumed to indicate the start of a postmodifying phrase.

Once triggered, the rules in Table 6.11 attempt to solve **Question 2** by investigating the following tokens, and excluding the postmodifier from consideration if found to

be non-restrictive. The study of postmodifiers in the pilot annotation study revealed that only four out of 60 superlatives (around 7%) had a postmodifier that was marked non-restrictive by the annotators. This implies that given a postmodified comparison set, there is a chance of approximately 93% that the postmodifier is restrictive. (The probability may actually be even higher since some comparison sets have more than one postmodifier.) Due to the small size of the data set, developing strategies for excluding non-restrictive postmodifiers represented a problem and was largely based on intuition. I found that the main types of postmodifiers that can be non-restrictive are found in the `ppPhrase` and `relClause` types.

The following examples illustrate how `CS-Identifier` presently tries to take non-restrictive postmodifiers into account. Once a token has been recognised as preposition, a specially created lexicon is consulted which lists `ppPhrase` exclusions and restrictions. Excluded are all instances of “although”, “as”, “because”, “despite”, “except”, and “that”, because they are unlikely to describe a PP that restricts the comparison set.<sup>1</sup> A number of other prepositions are listed in the lexicon with restrictions: For example, “at” is excluded from consideration if it is followed by a cardinal number (CD), as it is then likely to be non-restrictive. For example, in (6.9), the preposition “at” is followed by “53.9 km”, which specifies the length of the tunnel, but does not restrict the comparison set. Similarly, “with”-PPs are only included if the second token of the phrase is not tagged IN and is not an indefinite determiner (*a/an*), which (correctly) excludes “with about one-fifth world share” in (6.10), but includes “with only the right hand” in (6.11). (In all examples, the focus token is highlighted in bold font). Further examples can be found in Appendix D.1.

(6.9) The Seikan Tunnel in Japan is the *longest* [rail tunnel]<sub>{CSHead}</sub> in the world] **at**<sub>{IN}</sub> 53.9<sub>{CD}</sub> km (33.4 miles), of which 23.3 km (14.5 miles) is under the sea.

[Tunnel]

(6.10) In 2005, Russia was the *largest* [producer]<sub>{CSHead}</sub> of nickel] **with**<sub>{IN}</sub> about<sub>{IN}</sub> one-fifth world share closely followed by Canada, Australia and Indonesia, as reported by the British Geological Survey.

---

<sup>1</sup>The token “that” is treated separately in the `relClause` module.

[Nickel]

- (6.11) The word “hypolimnion” is one of the *longest* [words<sub>{CSHead}</sub> that can be typed **with**<sub>{IN}</sub> only<sub>{RB}</sub> the right hand on a QWERTY and AZERTY keyboard].

[Hypolimnion]

While the lexicon approach is able to recognise some ppPhrase instances as non-restrictive, further work is needed to fine-tune the restrictions and exceptions, as some of them only offer partial solutions. For example, while a “measure” PP like (6.9) is successfully excluded, the method would fail to recognise it if the cardinal number was preceded by the token “around” (“at around 53.9 km”). More data will be needed to deal with such cases effectively. Given the relatively low frequency of non-restrictive postmodifiers, this is left for future work.

To solve the **Question 3**, any rule that has been triggered and found to be non-restrictive tries to match as many of the tokens following the token in focus (“focus token”) as possible. When a rule succeeds, all matching tokens are added to the CSMain list, and the next unsubsumed token is returned as new focus token. For example, for relClause to succeed, the focus token needs to be tagged WDT (*wh*-determiner), WP (*wh*-pronoun), or have the lemma “that” (which can also be tagged as IN). After adding the token to the CSMain list, the word tokens to the right are investigated, and appended to the list if they conform to a particular pattern, consisting mainly of NP, VP and ADVP chunks. Crucially, as a verb is an obligatory component of a relative clause, the first VP chunk encountered acts as a stopper. This means that all tokens to the left (and including the VP chunk) are appended to the CSSpan list, and any I-NP tokens directly following the VP are also added (6.12). If the Identifier encounters a preposition or adjectival chunk during processing, the relative clause is interrupted at that point, and the following tokens are recognised via ppPhrase or adjPhrase. ADVP chunks, however, are included in the pattern as they can occur before the VP chunk.<sup>2</sup>

- (6.12) A similar comparison has been drawn among Go, chess and backgammon, perhaps the three *oldest* [games<sub>{CSHead}</sub> **that**<sub>{that}</sub> still<sub>{I-ADVP}</sub> enjoy<sub>{I-VP}</sub>

<sup>2</sup>The module is not yet able to correctly identify nested relative clauses (such as “*The house that Peter who I know owns*”). However, cases like this are assumed to only occur very rarely in superlative NPs.

worldwide<sub>{I-NP}</sub> popularity<sub>{I-NP}</sub>].

[Go (board game)]

The following examples show some complex CS spans CS-Identifier is able recognise with the approach described in the previous paragraphs.

- (6.13) “Sizdah bedar” is the *oldest* [prank-tradition<sub>{CSHead}</sub> in the world<sub>{ppPhrase}</sub> still<sub>{advPhrase}</sub> alive<sub>{adjPhrase}</sub> today<sub>{advPhrase}</sub>], which has led many to believe that the origins of the April Fools Day goes back to this tradition which is believed to have been celebrated by Persians as far back as 536 BC.

[April Fools’ Day]

- (6.14) The word “hypolimnion” is one of the *longest* [words<sub>{CSHead}</sub> that can be typed<sub>{relClause}</sub> with only the right hand<sub>{ppPhrase}</sub> on a QWERTY and AZERTY keyboard<sub>{ppPhrase}</sub>].

[Hypolimnion]

The examples illustrate that the above approach can recognise complex CS spans. However, it is affected by two main types of problems: Problems which CS-Identifier can not yet deal with efficiently, and problems resulting from imperfect tagging/chunking. One of the problems that represent a problem to CS-Identifier was discussed above (distinguishing restrictive from non-restrictive postmodifiers). Another important problem is found in the issue of **coordination**, as illustrated by (6.15) and (6.16).

- (6.15) Elected in 1307 when he was only 22 years old, Baldwin was the *most important* [Archbishop<sub>{CSHead}</sub> **and** Prince-Elector<sub>{I-NP}</sub> of Trier in the Middle Ages].

[History of Trier]

- (6.16) It is one of Bavaria’s oldest and *most popular* [festivals<sub>{CSHead}</sub>] **and** it<sub>{PRP}</sub> has increased to one of the biggest events in Germany.

[Gäubodenvolksfest]

In (6.15), “and” describes an NP coordination between “Archbishop” and “Prince-Elector”, and should therefore be included in the CS span (along with the following NP). In (6.16), on the other hand, “and” represents a coordination of two main clauses, and should not be included. Similarly, commas can indicate NP coordination, as for example in “the best writer, director, and actor”. What makes such cases more difficult to identify than “and” coordination is the fact that commas also often indicate apposition (6.17).

- (6.17) The entire district contains 14 Israeli settlements, with a total population of 10,000 and two of the *largest* [Palestinian refugee camps in the West Bank], Askar **and** Balata, which comprise about 8% of the total district population.

[Nablus]

CS-Identifier includes two modules **andMethod** and **commaMethod** to deal with these issues (further described in Appendix D.1). However, coordination is generally known to be a difficult problem in NLP, and the modules are only able to provide partial solutions.

A second major problem in the approach is caused by **imperfect tagging/chunking**, which may be due to genuine errors on the one hand, and definitions that are not sufficient for the current task on the other. As already mentioned in previous chapters, the software used in the experiments, the C&C tools, has been developed to correctly deal with frequently occurring phenomena in its target text type, in order to achieve the highest possible performance score. Since the current data represents a text type different from the one the C&C tools have been developed for, and since sentences containing ISA-1 superlatives are relatively rare, the performance of the tools may be worse for these reasons. I encountered a variety of tagging errors, and tried to address them in the rules wherever possible (i.e. where it did not have a negative effect on the rule’s precision). One example where tagging errors were not addressed because doing so resulted in lower precision was in the rule for past-participial non-finite clauses (`nonfinClause_VBN`), which deals with cases like (6.18).

- (6.18) The *most common* [`types_{CSHead}` of `art_{ppPhrase}` **created** `{VBN}` using models] are figure drawing, figure painting, sculpture and photography.

[Model (art)]

To trigger the rule, the focus token needs to be tagged VBN (past participle verb form). However, the tagger frequently mistakes such cases as the past tense form (VBD), in which case the rule fails. If the rule also took the tag VBD into account, it would result in a large number of False Positives (where VBD is in fact a past tense verb), thus lowering precision.

In addition to tagging errors, I also found that the output of some tools was not suitable for solving particular tasks. For example, while the chunker includes labels for PP chunks (B-PP and I-PP), these are no more suitable for identifying prepositional phrases than POS tags are. As the chunker has been trained to focus on noun phrases (NPs), in an example like “in the house”, the chunker usually only labels the preposition as I-PP (or, rarely, B-PP), but treats the rest of the PP as an NP chunk: “*in*/I-PP *the*/I-NP *UK*/I-NP”. A further problem affecting the `ppPhrase` module is that the POS tag IN can stand for both prepositions and subordinating conjunctions like “because” or “although”. This particularly represents a problem in cases where a token is ambiguous between a preposition and a subordinating conjunction, as illustrated by *since* in examples (D.4) and (D.5) in the Appendix. Again, `CS-Identifier` is only able to offer partial solutions to issues like this.

#### 6.2.4 Identifying the target span (T-Identifier)

The second component of `T/CS-Identifier` deals with the recognition of the target span of ISA-1 superlatives. It has to solve two major tasks: First of all, the target phrase has to be located in the sentence. The second task is then to identify its exact span, including all relevant pre- and postmodifiers as well as its determinative phrase.

To solve the task of locating the target phrase in the sentence, `T-Identifier` first checks the output of the four methods used to recognise ISA-1 superlatives, (described in Chapter 5.3.3.7), as these generally only succeed if a potential target item is found (although there are some minor rules where identification of a target is not necessary). This item is then marked up as target head (“`THead`”). If there are several potential target items, which is often the case with appositions, `T-Identifier` chooses the one closest to `CSHead`.

For cases where `ISA1-Identifier` was unable to detect a target, `T-Identifier` uses a back-off strategy to detect the target phrase. First, it searches the sentence for in-

stances of the verb “to be”, and if successful chooses the one that is nearest to CSHead in the sentence (but excluding instances that follow *wh*-determiners, which introduce subordinate clauses). It then distinguishes between two cases. If the index of the chosen verb is smaller than the index of CSHead, the method labels the first token to the left of “be” as THead (6.19). If the token is a comma (or closing bracket), the index is moved further left until a second comma (or opening bracket) is encountered, and the first token to its left is labelled THead (6.20). Otherwise, if the index of “to be” is greater than the index of CSHead, it chooses the first token to its right and using the same procedure for commas as just described.

(6.19) The population was 193,830 in 1999 (census), and it<sub>{THead}</sub> **is** currently the fastest growing town in Kenya, and currently the 5th *largest*<sub>{CSHead}</sub> [in Kenya].

[Eldoret]

(6.20) Forum Hadriani<sub>{THead}</sub>, at the modern town of Voorburg, **was** the northernmost Roman city on the European continent and the second *oldest* [city<sub>{CSHead}</sub> of The Netherlands].

[Forum Hadriani]

Once a potential THead is found, T-Identifier attempts to determine the span of the nominal phrase the target is assumed to occur in. T-Identifier first investigates the context to the left of THead by using the modules shown in Table 6.2. Generally, they work in a similar way to the PostMod modules described in the previous section, but are applied backwards (moving from right to left). The module npPhrase\_PreMod aims to identify NP sequences, while adjPhrase\_PreMod and advPhrase\_PreMod look for adjectival and adverbial phrases. The goal of the detPhrase\_PreMod method is to locate the determiner of the target, and andMethod\_PreMod contains a set of rules to deal with coordination. T-Identifier does not apply PreMod rules if THead has one of the tags VB, VBG, IN, or TO, or the lemma “that”, as these indicate the start of the target phrase (6.21).

(6.21) Being<sub>{VBG}</sub> raised to the Blood, while exceptionally rare, is the *greatest* honor possible for one who was born a commoner.

[Seanchan]

T-Identifier then considers the context to the right of THead, using the postmodifier modules developed for the identification of the CSSpan. Some of the rules are slightly modified: For example, the relClause and relClause\_BNP methods fail if the word preceding the focus token has a Named Entity tag, as Named Entity targets usually do not require defining relative clauses. Generally, cases where THead was determined via the back-off method are at a disadvantage. While THead is assumed to be the head of the target phrase, this is not necessarily the case for back-off targets. This particularly applies to cases which were selected because they occur to the right (rather than to the left) of the verb “to be”: Since English NPs are right-headed, the back-off target head is likely to be a determiner or premodifier rather than the head (e.g. “...is the<sub>{THead}</sub> new cathedral in the centre.”). Such cases can still be identified correctly via the PostMod rules, but it is of importance to note that their internal structure is not analysed correctly.

Type	Module
I.	npPhrase_PreMod
II.	adjPhrase_PreMod
III.	adjPhrase_PreMod
IV.	detPhrase_PreMod
i.	andMethod_PreMod

Table 6.2: Modules recognising THead premodifiers and determinative phrase

## 6.2.5 Results and discussion

T/CS-Identifier is evaluated on two separate data sets. First, its performance is tested on the set of ISA-1 superlatives in the TextWiki test set, with separate figures for the set of gold-standard ISA-1 superlatives, and the subgroup of ISA-1 superlatives that were detected by SUP-Classifer in the previous experiment. Secondly, the performance of T/CS-Identifier is also evaluated on the first 100 sentences in Bos and Nissim (2006)’s WSJ test data set, and compared to the results of their DLA system on the same set of sentences. The second evaluation is described in Section 6.3.

### 6.2.5.1 Baseline systems

Bos and Nissim (2006) developed two baseline systems for comparison set determination, which I recreated according to their instructions, however with some slight modifications. For Baseline 1, Bos and Nissim take the first word following the superlative as the beginning of the comparison set, and the first word tagged as NN.\* in that sequence as the end. Their second baseline takes the first word after the superlative as the beginning of the comparison set, and the end of the sentence as the end (excluding the final punctuation mark). They note that this approach is likely to generate comparison sets much wider than required. This baseline, however, can be improved easily by not just using the end of the sentence as break-off point, but the nearest full-stop or comma. For a fair evaluation of the results, I have decided to amend Bos and Nissim’s second baseline system accordingly.

Since the current CS-Identifier also takes into account the determinative phrase of the superlative NP, both of the above baseline systems check if the token preceding B-SUP is “the” (which has been shown to be the most frequent one), and if successful, return it as determiner. The CS span is marked as correct only if both its components CDet and CMain are exact matches with the gold standard. Finally, the baseline system for T-Identifier identifies the sequence of I-NP chunks closest to the superlative (B-SUP) as target span. All baseline results are shown in Table 6.3.

Baseline	Accuracy
CS-Baseline 1	12.4% (11/89)
- CSDet	70.8% (63/89)
- CSMain	28.1% (25/89)
CS-Baseline 2	27.0% (24/89)
- CSDet	70.8% (63/89)
- CSMain	46.1% (41/89)
T-Baseline	16.9% (15/89)

Table 6.3: Baseline results for gold-standard ISA-1 superlatives in TextWiki test set (Accuracy)

### 6.2.5.2 Evaluation on the TextWiki test set

Table 6.4 shows the performance of T/CS-Identifier and its components on the gold-standard TextWiki test set, measured in accuracy. All components clearly outperform their respective baselines, with particularly good results for CS-Identifier (88.8% accuracy). T-Identifier does moderately well with 58.4%. As will be further explained below, its lower performance is mainly due to the fact that it needs to solve two tasks: Locating the target head in the sentence, and identifying its pre- and postmodifiers. Due to its lower results, the overall result for T/CS-Identifier (in terms of correctly identified T and CS spans) lies only at 52.8%.

System	Accuracy
CS-Identifier	88.8% (79/89)
- CSDet	98.9% (88/89)
- CSMain	89.9% (80/89)
T-Identifier	58.4% (52/89)
T/CS-Identifier	52.8% (47/89)

Table 6.4: Performance of T/CS-Identifier and its components on gold-standard ISA-1 superlatives in TextWiki test set (Accuracy)

When T/CS-Identifier and its components are evaluated on the subset of ISA-1 superlatives that were identified by SUP-Classifier in previous experiment, the overall accuracy result for T/CS-Identifier improves considerably by around 10%, as shown in Table 6.5.

System	Accuracy
CS-Identifier	90.7% (68/75)
- CSDet	98.7% (74/75)
- CSMain	92.0% (69/75)
T-Identifier	69.3% (52/75)
T/CS-Identifier	62.7% (47/75)

Table 6.5: Performance of T/CS-Identifier and its components on ISA-1 superlatives identified by SUP-Classifier in TextWiki test set (Accuracy)

All of the sentences can be found in Appendix C.3 along with their gold standard

annotation and T/CS-Identifier results.

The ten errors caused by CS-Identifier are due to one mistake in CSDet span determination, and nine mistakes in CSMain span determination. The CSDet span that was incorrectly recognised is shown in (6.22).<sup>3</sup>

- (6.22) {Secret of Mana} is also [the number 6] *most remixed* [soundtrack on the popular video game music site OverClocked ReMix], with Seiken Densetsu 3 tied at 18.

[*Mana (series)*]

Clearly, this is an unconventional type of determinative phrase in which the string “number 6” expresses an ordinal position in the comparison set (similar to “6th”). As no examples like this were encountered in the development set, the pattern shown in Table 6.1 does not yet take them into consideration.

Three of the nine errors in the CSMain span determination were caused by the modules recognising non-finite clauses (Type IV in Table 6.11). The nonfinClause\_VBN module failed to identify both (6.23) and (6.24) due to tagging errors: The focus words *encountered* and *found* were both incorrectly tagged as past tense verb forms (VBD) instead of past participle forms (VBN). In (6.25), on the other hand, the error was caused by a restriction in the gerund-participial clause module (nonfinClause\_VBG), where focus words with lemma “to be” are excluded from consideration, as they are more likely to express the ISA relation between target and comparison set (as shown in example (D.10) in the Appendix).

- (6.23) {AMS} is [the] *most frequent* [type]<sub>{CSHead}</sub> of altitude sickness<sub>{ppPhrase}</sub> encountered].

[*Altitude sickness*]

- (6.24) [The] *most common* [organisms]<sub>{CSHead}</sub> **found**] are {Campylobacter (from animal products), Salmonella (also often from animal foodstuffs), Cryptosporidium (ditto), and Giardia lamblia (lives in water)}.

<sup>3</sup>Gold standard target spans are indicated by curly brackets, and gold standard CSDet and CSMain spans by square brackets. The prediction of T/CS-Identifier is shown by use of underlined text.

[*Infectious diarrhea*]

- (6.25) [The] *most common* [cause]<sub>{CSHead}</sub> of a person<sub>{ppPhrase}</sub> **being** underweight] is primarily {malnutrition caused by the unavailability of adequate food}, which can run as high as 50% in parts of sub-Saharan Africa and south Asia.

[*Underweight*]

Two further errors were caused by the *advPhrase* module, both of which concern the same comparison set, shown in (6.26). The word *ever* is tagged RB (adverb), but occurs within a VP chunk (I-VP). Since the module only focuses on adverbial chunk tags and NE tags, it fails to recognise *ever* as adverb. This should be taken into account in an updated version of T/CS-Identifier.

- (6.26) {The Human Genome Project} - [the] *largest*, most costly [single biological study]<sub>{CSHead}</sub> **ever** undertaken] - began in 1988 under the leadership of James D. Watson, after preliminary work with genetically simpler model organisms such as *E. coli*, *S. cerevisiae* and *C. elegans*.

[*History of biology*]

The remaining four errors were caused by preprocessing modules and a bug in the *ppPhrase* method. The CSHeads of (6.27) and (6.28) were wrongly identified by the *findHead* method, and the error in (6.29) results from the method *addSuperlativeTag* falsely adjusting the I-SUP index (both methods were described in Section 5.3.3). The token “and” was falsely included in the comparison set span in (6.30) because of a bug in *ppPhrase*, which will be easy to fix in an updated version of the Identifier.

- (6.27) Starting with the largest, Mikhail Gorbachev , then Leonid Brezhnev (Yuri Andropov and Konstantin Chernenko almost never appear due to the short length of their respective terms), then Nikita Khrushchev, Josef Stalin and finally [the] *smallest*, {Vladimir Lenin}.

[*Matryoshka doll*]

- (6.28) When Baldwin IX, Count of Flanders and Hainault, left on the Fourth Crusade in 1202, he left his western domains under [his] *eldest* [daughter] {Joanna}.

[*War of the Succession of Flanders and Hainault*]

(6.29) {A gonk in computer-speak} is [the user's] *least favourite* [piece of hardware].

[*Gonk*]

(6.30) At 670 ha (1,656 acres) in extent, {Hirta} is [the] *largest* [island in the group] and comprises more than 78% of the land area of the archipelago.

[*St Kilda, Scotland*]

While T-Identifier clearly outperforms the baseline system in recognising target spans, its results are considerably lower than those of CS-Identifier. The reason for this is that it has to solve two difficult tasks: First of all, locating the target head (THead), and secondly identifying the target span. This is a more difficult task than identifying the CS span, because the target is not necessarily limited by a preceding determinative phrase marking the beginning of the span. In addition, the target phrase is not restricted to being a noun phrase.

Table 6.6 shows how T-Identifier's individual methods for locating the target head performed on the TextWiki test set. The four ISA1-Identifier methods, which are summarised in Table 5.17, retrieved a target head for 67 out of 89 ISA-1 superlatives. Only four of these have been found to be responsible for a target span not being identified correctly (6%). For example, in (6.31), Method Ib. (see Table below) identified "way" instead of "Rabite" as THead.

(6.31) {The Rabite} has become a sort of mascot for the Mana series, much the same way as the Chocobo represents Final Fantasy, and is [one of its] *most recognizable* [icons].

[*Mana (series)*]

The back-off method, on the other hand, did not perform well on the test data: 8 out of the 9 times that it was applied it was found to be directly responsible for T-Identifier failing to identify the correct target span. Since there is no back-off strategy for cases where T and CS stand in apposition, there are six superlative sentences altogether

which have been assigned no target at all.

Method	Total (89)	Cause of error (out of 37)
Ia. CS in <code>ncsubj</code> position	33	2
b. CS in <code>xcomp</code> position	31	1
IIa. T and CS in <code>ncsubj</code> apposition	2	0
IIb. T and CS in <code>dobj</code> apposition	6	1
Back-off method	9	8
Other	2	0
(No target found)	6	6

Table 6.6: Assessment of the target location methods

Table 6.6 shows that of 37 errors produced by `T-Identifier` altogether, around half (18/37) result from incorrect or failed target location (18 cases). While target location via `ISA1-Identifier` methods works reasonably well, more work needs to be done on a back-off method, which will not only involve improving the existing one, but also implementing a back-off strategy for `ISA-1` appositions.

The other 19 errors are caused by `T-Identifier`'s `PreMod` and `PostMod` strategies. A closer investigation of the data shows that the problems fall into a number of different categories, shown in Table 6.7.

Error cause	Total	Examples in Appendix C.3
Brackets	8	(2), (3), (26), (34), (58), (85), (88), (89)
Coordination	5	(47), (55), (56), (62), (64)
Colon	1	(33)
<code>PreMod</code> modules	3	(35), (37), (43)
<code>PostMod</code> modules	1	(7)
Tagging errors	1	(40)

Table 6.7: Types of errors caused by `PreMod` and `PostMod` strategies

The most common type of error was caused by brackets within the target span, which have been excluded from consideration in the pre- and postmodifier rules. While they often introduce additional information that is not vital for inclusion in the target span, they occur frequently within the target span, which is interrupted if they are not taken

into account (6.32). Future version of T/CS-Identifier should include a further module that is able to handle brackets like the ones shown below.

- (6.32) [The] *most effective* [source of supplemental oxygen at high altitude] are {oxygen concentrators that use vacuum swing adsorption (VSA) technology}.

[*Altitude sickness*]

The second most common type of error is a result of coordination ambiguities, as for example in (6.33). Here, the target span should only consist of “Angel Falls”, while T-Identifier also included “the Amazon Basin”. Cases like these are difficult to deal with because the Identifier does not yet recognise whether the apposition relates to both parts of the coordination or not. One possible way of refining the system would be to take number information into account - if the CShead is in plural form, both parts are included, and if it is in singular form, only the closest part (in this case, “Angel Falls”), is included.

- (6.33) To the south, the dissected Guiana Highlands is home to the northern fringes of the Amazon Basin and {Angel Falls}, [the world’s] *highest* [waterfall].

[*Venezuela*]

The remaining errors caused by the PreMod and PostMod modules are mainly due to peculiarities of the tagsets that have not been considered during development on the one hand, and to unanticipated tagging errors on the other. For instance, “While” in (6.34) has the chunk tag I-NP, which causes the npPhrase.PreMod to falsely include it in the target span. This could be avoided by specifying that words tagged as prepositions (IN) should be excluded from the rule.

- (6.34) While {the arithmetic density} is [the] *most common* [way of measuring population density], several other methods have been developed which aim to provide a more accurate measure of population density over a specific area.

[*Population density*]

## 6.2.6 Conclusion

The analysis of the TestWiki test set results shows that T/CS-Identifier can still be further improved, in particular the component dealing with target strings. The CS-Identifier already represents a powerful tool for identifying CS strings and their determiners. The next section shows that it not only works well on ISA-1 superlatives, but that it also outperforms Bos and Nissim (2006)'s DLA tools on their WSJ data set.

## 6.3 Evaluation of the Superlative Relation Extractor (SRE) on Bos and Nissim (2006)'s data set

### 6.3.1 Overview

This section describes an evaluation of SUP-Classifier and T/CS-Identifier on the WSJ test set used by Bos and Nissim (2006) in their experiments. The motivation for this is not only a comparison of the performance of T/CS-Identifier with Bos and Nissim's DLA system, but also an investigation of the performance of the system on newspaper text.

The set consists of 217 sentences<sup>4</sup>, which were annotated with class labels according to the guidelines in Chapter 4.2.1. As Bos and Nissim have only included attributive superlatives in their test set, it is important to bear in mind that the distribution of superlatives is not representative of the genre as a whole. As some of the classes have been defined to not occur in an NP, I expect them to occur with zero frequency (FREE and PP). The results of running the classifier on the data set can be found in Section 6.3.2.

In addition, the first 100 sentences in the data set have also been annotated with T and CS spans according to the guidelines in Chapter 4.2.2. Testing the T/CS-Identifier on this data is interesting in several ways. First of all, its performance on newspaper data can be assessed. While the Identifier has been developed on a different genre (encyclopedia texts), it strongly relies on the C&C tools, which were developed on the WSJ corpus (Clark and Curran, 2004b). One would therefore expect less errors in their tagging/parsing performance on data, which in turn is likely to have a positive effect

---

<sup>4</sup>Cases where their data indicates that the parser failed were excluded from consideration.

on the performance of the Identifier, as one of its main error sources was in the task of locating the target head. However, as previously mentioned, one needs bear in mind that statistical parsers are tuned to produce correct results on *frequent* phenomena, no matter what the consequences are for constructions that occur relatively infrequently (as is the case for many superlative constructions). The results of the evaluation can be found in Section 6.3.3.

### 6.3.2 SUP-Classifier

Table 5.21 shows the result of automatic classification of the attributive superlatives in Bos and Nissim’s WSJ data set. As expected, there are no gold standard PP instances in the set. There is, however, one instance of the FREE class, which should probably have been excluded from the study, as it is clearly not attributive (6.35). There are no instances of the PROP class, which shows that they must have been excluded from consideration in the manual annotation.

- (6.35) Third and *most* important , Amex would charge me a far higher premium than other reputable companies would on a straight term policy for the same amount ; [*sic*]

Class	Precision	Recall	F-Measure
<b>IDIOM</b>	<b>88.9%</b> (32/36)	<b>100%</b> (32/32)	<b>94.1%</b>
<b>PP</b>	- (0/0)	- (0/0)	-
<b>PROP</b>	- (0/2)	- (0/0)	-
<b>ADV</b>	<b>66.7%</b> (4/6)	<b>40.0%</b> (4/10)	<b>50.0%</b>
<b>FREE</b>	<b>0%</b> (0/3)	<b>0%</b> (0/1)	<b>0%</b>
<b>INDEF</b>	<b>50.0%</b> (1/2)	<b>50.0%</b> (1/2)	<b>50.0%</b>
<b>ISA-1</b>	<b>89.4%</b> (59/66)	<b>84.3%</b> (59/70)	<b>86.8%</b>
<b>ISA-2</b>	<b>80.0%</b> (8/10)	<b>61.5%</b> (8/13)	<b>69.5%</b>
<b>DEF</b>	<b>89.1%</b> (82/92)	<b>92.1%</b> (82/89)	<b>90.6%</b>

Table 6.8: SUP-Classifier results

The high frequency of the IDIOM class is largely due to the word “latest”, which is considered idiomatic (cf. Chapter 3.2.4.1). Furthermore, the fact that there are 10 gold

standard instances of the ADV class shows that Bos and Nissim’s definition of their adverbial class differs from the definition used by the present study: They classify cases like (6.36) as attributive rather than adverbial, while the classification described in Chapter 3.2.3 includes them in the ADV class, on the grounds that “used” represents an action rather than a property inherent to the members of the comparison set. Bos and Nissim, on the other hand, only consider the superlative’s position with respect to the NP (which in this case is attributive).

- (6.36) This is probably the *most* widely used order – and the one most open to abuse by unscrupulous floor brokers , since it imposes no price restrictions .

The classes which correspond to Bos and Nissim’s attributive type of superlative are INDEF (with only two instances), ISA-1, ISA-2, and DEF. SUP-Classifier performed remarkably well on the ISA-1, ISA-2, and DEF classes, especially compared with their results on the TextWiki test set (Table 5.21). Both ISA-1 and DEF have better F-measure values (3.5% and 13.9% higher than in TextWiki, respectively). In the case of ISA-1, this is mainly due to a higher precision value (89.4% vs. 82.4% in TextWiki), which is likely to be due to a better performance of the parser as it was trained on WSJ data. ISA-1’s recall value, on the other hand, is lower on the current data. This is mainly due to the large proportion of appositions in newspaper language, which are harder to detect for the Identifier. The DEF class, on the other hand, shows an improvement of both precision and recall values compared to the TextWiki test set (89.1% vs. 69.7% precision, and 92.1% vs. 85.2% recall), which is a direct result of the higher precision values of the components earlier on in the pipeline.

### 6.3.3 T/CS-Identifier

To assess the performance of T/CS-Identifier on Bos and Nissim’s data, it was run twice on the first 100 sentences of their test set: First, only superlatives that were annotated as ISA-1 were included (40 instances altogether), which allows an assessment of all components of the Identifier (T, CSDet and CSMain) on newspaper data. Then, the Identifier was applied to the whole set of 100 sentences to investigate how CS-Identifier performs when *all* superlative classes are taken into account (not just ISA-1). These results are compared with the performance of Bos and Nissim’s system (DLA 4) on the same data.

Column 2 in Table 6.9 shows the results of `T/CS-Identifier` tested on the 40 gold-standard ISA-1 superlatives in the data set. Compared with its performance on the `TextWiki` test set (shown in Table 6.4), results have improved for all components. In particular, `T-Identifier` was able to correctly identify 72.5% of all target strings, an improvement of almost 15%. Again, this is likely to be the result of a better performance of the parser.

System	Accuracy (ISA-1)	Accuracy (all classes)
CS-Identifier	87.5% (35/40)	84.7% (83/98)
- CSDet	100% (40/40)	100% (98/98)
- CSMain	87.5% (35/40)	84.7% (83/98)
T-Identifier	72.5% (29/40)	-
T/CS-Identifier	67.5% (27/40)	-
DLA 4	77.5% (31/40)	72.4% (71/98)

Table 6.9: Performance of `T/CS-Identifier` and Bos and Nissim (2006)'s DLA system on WSJ test set (Accuracy)

A closer look at the data further supports this. Table D.2 in Appendix C.3 shows a list of all sentences whose target span was not identified correctly by `T-Identifier`. Two of them (5%) have not been assigned a target head, which is comparable to the results on the `TextWiki` data set, where six out of 89 ISA-1 superlatives were not assigned a head (6.7%). An investigation of the sentences in Table D.2 shows that, unlike in the `TextWiki` evaluation, `T-Identifier` did well in locating the target head (`THead`): The back-off method only had to be used twice altogether, and none of the ISA1-Identifier methods were responsible for `T-Identifier`'s errors (Table 6.10). As can be seen in Table D.2, all `THeads` identified by Methods I and II are valid target heads that occur within the gold-standard target span. This means that all errors (apart from the ones where `THead` was not correctly located) were caused by the Identifier's `PreMod` and `PostMod` methods. Table D.2 shows that the problems encountered are similar to the ones discussed in the `TextWiki` evaluation above.

With 87.5% accuracy, `CS-Identifier` has very good results on the ISA-1 subset. All determinative phrases were identified, and only five `CSMain` spans were not recognised correctly. Strikingly, `CS-Identifier` performs almost equally well when tested on the whole set of sentences (84.7%, see Column 2 in Table 6.9). Although it was originally

Method	Total (40)	Cause of error
Ia. CS in <i>ncsubj</i> position	4	0
b. CS in <i>xcomp</i> position	10	0
IIa. T and CS in <i>ncsubj</i> apposition	6	0
IIb. T and CS in <i>dobj</i> apposition	15	0
Back-off method	2	2
Other	1	0
(No target found)	2	2

Table 6.10: Assessment of the target location methods in WSJ

developed to only handle ISA-1 comparison sets, the results show that it can be reliably used to identify the comparison set span of *any* attributive superlative.

Finally, CS-Identifier’s results were compared with the results of Bos and Nissim’s DLA 4 system on the same set of sentences. A comparison of the results in Table 6.9 shows that in terms of accuracy on the complete set of sentences, CS-Identifier outperforms DLA 4 by around 12% (84.7% vs. 72.4%). As the two annotation procedures differ somewhat, especially with respect to the span of the determinative phrase, results are counted as correct if they match their respective gold standard annotation. As Bos and Nissim only mark up possessive determiners, while the current study has to identify the whole span (as described in Section 6.2.3.1), this clearly does not put Bos and Nissim’s system at a disadvantage in the evaluation. There are three cases where Bos and Nissim’s gold standard annotation for the *CSMain* span differs from the current guidelines. In (6.37), Bos and Nissim’s *CSMain* span excludes the substring “of how many plants could be in jeopardy”, which the present study views as NP-head complement to “indication”, and therefore a crucial constituent of the CS string (Huddleston and Pullum, 2002). In (6.38), Bos and Nissim have marked up “two” as belonging to the superlative span, comparable to “second largest”. The present approach treats “two” as postdeterminer and therefore part of the *CSDet* phrase. Finally, in (6.39), the substring “as measured by the Nasdaq financial index” is excluded from the CS span in Bos and Nissim’s approach, but included in this study’s gold standard, as it is considered to define the *CSHead* “issues”. In each of these cases, the systems were evaluated according to their own gold standard.

(6.37) But Mr. Iacocca ’s remarks are [the] *most specific* [indication to date of how

many plants could be in jeopardy] .

- (6.38) The spokeswoman said the move is n't directly a response to Quotron 's loss of [its two] *biggest* [customers] , Merrill Lynch & Co. and American Express Co. 's Shearson Lehman Hutton Inc. , to Automated Data Processing Inc. earlier this year .
- (6.39) [The] *largest* [financial issues , as measured by the Nasdaq financial index] , tumbled 3.23 to [sic]

Some comparison sets were not recognised correctly by either method (9 cases altogether). However, the systems generally failed for different reasons. For example, while neither of them managed to include the PP starting with “since” in (6.40), DLA 4's CSMain span cuts off after the token “leader”, while CS-Identifier recognises that the “since” phrase should be included. However, because “since” introduces a subordinate clause in this case (and not a PP), the ppPhrase methods fails on encountering the verb “began”.

- (6.40) At the same time , a recent poll shows that Mrs. Thatcher has hit the *lowest* [popularity rating of any British leader since polling began 50 years ago] .

There are 6 cases which were correctly identified by DLA 4, but where CS-Identifier produced an error. The errors are mainly of a similar kind to the ones described in Table 6.7. In one case, the PostMod “and” method incorrectly marked up a conjunction as belonging to the CS span (6.41). This example shows that in some cases semantic knowledge is vital to determine the span correctly.

- (6.41) Anheuser , [the world 's] *largest* [brewer] and U.S. market leader , has historically been reluctant to engage in price-cutting as a means of boosting sales volume .

CS-Identifier was able to find the comparison set span of a total of 18 cases where DLA 4 failed. Some types of error occur more than once, indicating a number of general problems in the approach. For example, it seems that the DLA system is not

able to handle coordination phenomena as in (6.42), where the possessive determiner is detached from the superlative or `CSMain` string. For both *largest* and *most successful*, DLA 4 fails to identify “the world’s” as possessive determiner. `CS-Identifier` handles cases like these via the back-off method described in Section 6.2.3.1.

(6.42) Like Mr. Geffen ’s arrangement , the venture gives Mr. Azoff a link to [the world ’s] *largest* and *most successful* [record distributor] ; in the U.S. alone , Warner has a 40 % share of the market , about double its nearest competitor , Sony Corp. ’s CBS Records .

Another problem DLA is unable to handle are colons, as shown in (6.43). Instead of marking up the last element before the colon as the end of the CS span, the predicted span stretches to the end of the sentence. `CS-Identifier`, on the other hand, recognises “in a larger design” as PP, and then stops because the new focus token “:” does not trigger any of the `PostMod` methods.

(6.43) But the transaction is just [Mr. Peladeau ’s] *latest* [step in a larger design] : to build Quebecor through acquisitions into an integrated paper , publishing and printing concern with a reach throughout North America .

A further common problem is illustrated in (6.44), where DLA failed to include the substring “yet in the renewed outcry against program trading” in the CS span. Generally, it seems that once the system encounters adverbial or adjectival chunks, it marks the previous token as the end of the span. Since `CS-Identifier` has rules for both adjectival and adverbial phrases, it is able to recognise the adverb “yet” via the `advPhrase` method, and then identifies “in the renewed outcry” and “against program trading” by using the `ppPhrase` method, resulting in the correct CS span.

(6.44) The move is [the] *biggest* [salvo yet in the renewed outcry against program trading] , with Kemper putting its money – the millions of dollars in commissions it generates each year – where its mouth is .

To see all examples where DLA fails but `CS-Identifier` manages to solve them, please refer to the table in Appendix D.3.

### 6.3.4 Conclusion

The evaluation of `SUP-Classifier` and `T/CS-Identifier` on Bos and Nissim (2006)'s data set provides some interesting results. First of all, compared to the results on the `TextWiki` test set, both components show improvement due to better performance of the parser on the data set it was developed on. However, the analysis also revealed that newspaper language is characterised by a large amount of appositions, which represent a problem to `ISA1-Identifier` in the classification, and `T-Identifier` in the target location task, particularly with respect to recall.

One of the striking results of this section is that `CS-Identifier` not only performs almost equally well on all superlative types in Bos and Nissim's data set (compared to just `ISA-1`, which it had been developed for), but it also outperforms Bos and Nissim's DLA 4 system on their own data, with 84.7% accuracy (compared to 72.4%). The error analysis also showed that on encountering new data it is possible to further refine the `PreMod` and `PostMod` methods of `T/CS-Identifier`. Improving the system with these new findings is left to future work.

Type	Module	Main triggers	Example CSMain strings (module output underlined)
I.	ppPhrase	IN, B-PP, I-PP	<i>substance <u>on earth</u></i>
II.	advPhrase	I-ADVP, B-ADVP, I-DAT	<i>Mayan astrological observatory <u>still intact</u></i>
III.	adjPhrase	JJ, B-ADJP, I-ADJP	<i>Mayan astrological observatory <u>still intact</u></i>
IV a.	nonfinClause_VBN	VBN	<i>composition <u>considered opera,</u> meat <u>cooked in master stock</u></i>
b.	nonfinClause_VBG	VBG	<i>idea <u>underpinning</u> <u>Hegelian Dialectics</u></i>
c.	nonfinClause_TOVB	TO+VB	<i>scripted series <u>to do so on a</u> <u>regular basis, Arabic author</u> <u>to mention the founding myth</u> <u>of Romulus and Remus</u></i>
V a.	relClause	WDT, WP, “that”	<i>number <u>that the</u> <u>eye can resolve as stripes,</u> <u>speed which allowed for</u> <u>sufficient sound quality</u></i>
b.	relClause_BNP	B-NP	<i>minority <u>you know about</u></i>
VI.	npPhrase	I-NP, POS	<i>park</i>

Table 6.11: Modules recognising CSHead postmodifiers

# Chapter 7

## Superlatives in Sentiment Analysis and Opinion Mining

### 7.1 Overview

In recent years, the domain of product reviews has attracted much attention in the area of sentiment analysis and opinion extraction. Being able to distinguish subjective from objective statements about products automatically is of great use not only for companies who want information about customer's opinions on their products, but also for recommendation systems whose purpose is to assist customers in deciding which product to buy.

The area of *Sentiment Analysis* is concerned with deciding whether a given document (or paragraph, sentence, phrase, or word) is subjective or objective, where subjectivity in natural language is usually taken to refer to “aspects of language used to express opinions and evaluations”. It furthermore aims to solve the problem of *polarity identification*, which has as its goal to identify the polarity of subjective elements, usually in terms of *positive*, *negative* and *neutral*. Most approaches to sentiment detection are based on supervised or semi-supervised machine learning techniques (e.g. Pang et al. (2002), Yu and Hatzivassiloglou (2003), Wilson et al. (2004), Kim and Hovy (2004)).

*Opinion mining* takes this one step further. Most of its studies focus on customer reviews on the web, such as product reviews (e.g. from Amazon) or book/movie reviews. Rather than only classifying documents or sentences as positive or negative, the interest here lies in extracting information about which entities or features of entities

are considered as positive or negative, and there have also been attempts at summarising this information (Hu and Liu (2004), Popescu and Etzioni (2005); Carenini et al. (2005)).

This chapter aims to present a first investigation of how the findings of this thesis could benefit applications in the area of Opinion Mining. In particular, it investigates how SUP-Classifier and T/CS-Identifier can be used to find product features in customer reviews, which is a major task in Opinion Mining. As well as demonstrating a practical application of the work described in this thesis, the chapter also examines the performance of the SRE system on an orally-oriented register (customer reviews).

## 7.2 Previous work and motivation

Opinion Mining systems are required to solve the following main tasks (e.g. Hu and Liu (2004)):

1. Feature Identification
2. Opinion Word Identification
3. Sentiment Classification
4. Opinion Summarisation

As a first step, the system tries to identify features of the products that customers are interested in, usually by using data mining and natural language processing techniques. Hu and Liu (2004) define the term “product feature” as representing both components of an object (e.g. zoom) and their respective attributes (e.g. size).<sup>1</sup> Then, the system needs to identify sentences in the reviews that express opinions about these features. This involves identifying which words indicate statements of opinions versus words that indicate statements of facts (*subjectivity recognition*). In the third step, the system has to determine whether a statement of opinion is positive or negative (*polarity identification*). Finally, the system also requires techniques for summarising the information gathered in the previous steps.

The focus of the current chapter is to show how Opinion Mining can benefit from the study of superlatives described in this thesis. In particular, I aim to show that it can

---

<sup>1</sup>In this definition, the object itself is also a feature.

help to solve the first two tasks, which represent crucial components in any Opinion Mining system. Existing work on identifying product features (Task 1) often relies on the simple heuristic that explicit features are expressed as noun phrases. While this narrows down the set of product feature candidates, it is clear that not all noun phrases represent product features. Various approaches to further limiting this set have been proposed. The two most notable ones are by Hu and Liu (2004) and Popescu and Etzioni (2005).

Hu and Liu (2004) suggest that nouns or noun phrases that occur frequently in reviews for a particular product are likely to be features. To identify frequent features they use association mining, and then apply heuristic-guided pruning to further refine their results. They also found that they can improve the F-measure of their system by assuming that adjectives appearing in the same sentence as frequent features are opinion words, thereby solving Task 2 (however, at the cost of precision). In particular, by retrieving nouns and noun phrases that co-occur with these opinion words in other sentences, this helps their system to identify so-called *infrequent features*, which have been shown to also be of great interest (Pang and Lee, 2008).

Popescu and Etzioni (2005), on the other hand, consider product features to be concepts that stand in particular semantic relationships with the product (for example, a camera may have “properties” size, weight, etc., while the lens, the flash, etc. stand in a “part” relationship with the camera). Their strategy for identifying such features is to search for corresponding meronymy discriminators. This approach achieves better performance than that the one employed by Hu and Liu (2004), but it does not yet involve any sentiment analysis, and opinion words have to be identified in a second step.

So far, none of the studies in Sentiment Analysis or Opinion Mining have specifically looked at the role of superlatives in these areas. While it has been generally acknowledged that there is a positive correlation between subjectivity and the use of adjectives, there has not yet been a thorough investigation of superlatives in this context. Jindal and Liu (2006b)’s study of comparative sentence mining (which was reviewed in Chapter 2.2.1), is clearly of relevance here, as is a recent follow-up study by Ganapathibhotla and Liu (2008), which builds on Jindal and Liu’s findings and focuses on determining which of the extracted entities in a comparison are preferred by its author. However, the set of “features” they determine as part of their comparative relation vector ( $\langle \text{relationWord} \rangle$ ,  $\langle \text{features} \rangle$ ,  $\langle \text{entityS1} \rangle$ ,  $\langle \text{entityS2} \rangle$ ) do not necessarily

correspond to the “product features” in Hu and Liu’s annotation of the corpus. As Jindal and Liu (2006b) apply their vector approach to *every* gradable comparative in their corpus, this involves a large amount of “features” which are not simultaneously “product features”. As a consequence, their system is not suitable for the task of identifying product features. Unlike Jindal and Liu’s study, which uses the same treatment for all “gradable comparatives”, the current study takes different types into account, and suggests that a particular subclass of superlatives (namely, ISA superlatives) is especially useful in identifying product features. Before elaborating on this idea, the next section provides an overview of the data.

### 7.3 Data

The experiments described in this chapter are centred around Hu and Liu’s corpus of customer reviews, which was not only the basis of their own study of opinion feature mining (2004), but has also been used as a test set by other studies as well (e.g. Popescu and Etzioni (2005)). Sentences in this corpus have been manually annotated with information about product features (addressing Task 1 above). Each feature is taken to express an opinion, and is labelled as *positive* or *negative* in terms of values on a six-point scale, where [+3] and [+1] stand for the strongest positive and weakest positive opinions, respectively, and [-3] and [-1] stand for the strongest and weakest negative opinions.

Hu and Liu’s corpus contains 4259 sentences altogether, of which 1728 include at least one product feature (40.6%). The remaining sentences in the corpus either contain no product feature (2217 altogether, 52.1%), or describe a review title, in which case they have been excluded from annotation (314 instances, 7.4%). Table 7.1 shows the total number of sentences, title sentences (T), feature-containing sentences (F), and non-feature containing sentences (N) for each file in Hu and Liu’s dataset (five altogether). The number of superlatives found in each of the files is displayed in the final column (S), with a total of 230 superlatives in 4259 sentences, which means that there is around one superlative in every 18 sentences.

All 230 superlatives found in the corpus have been annotated with class labels according to the guidelines in Chapter 4.2.1. Table 7.2 compares the distribution of superlative classes across the `Hu&Liu` data set and the `TextWiki` test set. While it is

File	Sentences	#T	#F	#N	#S
Apex	839	99	345	395	35
Canon	642	45	238	359	62
Creative	1811	95	720	996	80
Nikon	380	34	160	186	16
Nokia	587	41	265	281	38
TOTAL	4259	314 (7.4%)	1728 (40.6%)	2217 (52.1%)	230

Table 7.1: Statistics of Hu and Liu's data set (2004)

not possible to compare absolute numbers due to the different sizes of the data sets, the table shows the overall distribution of classes in percent (%). Some classes are represented equally in both the *Hu&Liu* data set and the *TextWiki* test set, e.g. DEF with 19.1% and 20.6% respectively, and members of the FREE class are equally rare in both data sets at 1.3%. The other classes, however, show considerable variation in their distribution, which suggests that the different classes may have different purposes.

Class	Hu&Liu	TextWiki
IDIOM	4.9% (11)	2.8% (11)
PP	12.0% (27)	5.6% (22)
PROP	20.4% (46)	29.3% (115)
ADV	4.4% (10)	7.9% (31)
FREE	1.3% (3)	1.3% (5)
INDEF	6.7% (15)	1.0% (4)
ISA-1	27.6% (62)	22.6% (89)
ISA-2	3.6% (8)	8.9% (35)
DEF	19.1% (43)	20.6% (81)

Table 7.2: Distribution of superlatives in *Hu&Liu* and *TextWiki* test set

The first thing to note is that PROP and ISA-1 have swapped position. While PROP was the most frequent class in *TextWiki* with 29.3% and ISA-1 was the second most frequent class with 22.6%, the situation is now reversed: ISA-1 occupies the Number 1 frequency slot in *Hu&Liu* at 27.6%, and PROP comes second with only 20.4%. The high frequency of PROP superlatives in *TextWiki* can be explained by the fact that encyclopedia entries usually define classes, and proportional quantifiers are a good

way of describing properties that do not apply to all members of a class (but to *most* of them). The high frequency of ISA-1 superlatives in customer review data, on the other hand, suggests that they may be an important means of evaluating products. Interestingly, the proportion of ISA-2 superlatives is much lower, with 3.6% compared to 8.9% in TextWiki. This is likely to be due to the fact that customer reviews are often written in first person and tend to express a single person's point of view, which means that there is less need for passive constructions. Customers are more likely to write "I think that X is the best Y" than "X is thought to be the best Y" (cf. Chapter 3.2.2.1).

Another striking difference in frequency can be found in the INDEF class. While only 1.0% of all superlative forms in TextWiki are INDEF members, their frequency is considerably higher in the corpus of customer reviews with 6.7%. The reason for this difference may be that customer reviews often display what can be referred to as "abbreviated language", which is characterised by the omission of certain linguistic elements, including definite articles (Janoschka, 2004). A closer look at the data shows that a great proportion of INDEF instances occur in review titles (e.g. "*best* non-slr digital camera under 5 megapixels"). The customer review data also has a higher proportion of PP superlatives, with a frequency of 12.0% compared to only 5.6% in TextWiki (which is likely to be due to an increased use of the phrase "at least").

## 7.4 Superlatives in Opinion Mining

This section aims to show that superlatives are special indicators of product features in customer reviews. In particular, I claim that this only applies to a subgroup of superlatives (ISA), and support this by presenting an investigation of the distribution of features across the nine superlative classes.

Table 7.3 shows the proportion of title sentences (T), feature-containing sentences (F), and non-feature containing sentences (N) among the 230 sentences that contain superlatives. The last row indicates that the proportion of feature-containing sentences among them is higher (at 51.7%) than the average for all sentences (40.6%) (last row in Table 7.1). What is striking, however, is that features are particularly highly represented among the ISA-1 and ISA-2 classes: Of 63 ISA-1 superlatives in the data set, 46 occur in a sentence involving a feature (73.0%), and of 8 sentences with an ISA-2

superlative 7 contain a feature (87.5%).<sup>2</sup> This suggests that membership in the ISA class is a good indicator of the sentence containing a product feature.

Class	#S	#T	#F	#N
IDIOM	12	0	5	7
PP	27	1	13	13
PROP	47	0	23	24
ADV	10	0	4	6
FREE	3	1	2	0
INDEF	15	10	3	2
ISA-1	63	2	46	15
ISA-2	8	0	7	1
DEF	45	9	16	20
TOTAL	230 (100%)	23 (10%)	119 (51.7%)	88 (38.3%)

Table 7.3: Distribution of features

A closer investigation of the superlative instances in the data set reveals further interesting results. Among the 119 superlative sentences that contain a feature (column “F”), not all superlatives directly contribute to the evaluation of the feature. For example, the superlative “most” in (7.1), which belongs to the PROP class, is not directly involved in the evaluation of the feature “firewire” as [-1]. In contrast, the ISA superlative “best” in (7.2) is directly responsible for the positive [+3] rating of the feature “dvd player”.

(7.1) it does n’t have firewire , not a real complaint since *most* windows users do n’t generally have firewire cards themselves .

(7.2) i think , apex is the *best* dvd player you can get for the price .

An assessment of all feature-containing sentences with respect to the involvement of the superlative in the feature-rating shows that the IDIOM, PP, and PROP classes are of little relevance, while ISA-1 and ISA-2 clearly are, with the superlative form acting as opinion word evaluating the feature, or acting as intensifier of an opinion word, as

<sup>2</sup>All gold-standard ISA-1 superlatives in the data set are shown in Table E.2 in the Appendix.

for example “complaint” in (7.3). Furthermore, in 34 out of the 46 instances (73.9%), the feature is a substring of either the target or the comparison set.

(7.3) [my] *biggest* [complaint] is {the battery life or lack there of} .

These findings suggest that ISA-1 superlatives can be seen as special indicators of product features: They simultaneously solve Opinion Mining tasks 1 and 2 (see above) by containing a product feature within their T or CS string, as well as by expressing its respective opinion word.

An investigation of the 15 ISA-1-containing sentences in the data set that did not receive a feature label in Hu and Lu’s annotation (column “N” in Table 7.3) shows that some of them do in fact modify a feature. For example, (7.4) and (7.5) make a similar positive statement about a camera, however only (7.4) was annotated with a feature (player[+3]). To be consistent, (7.5) should receive the same feature label. Examples (7.6), on the other hand, is similar to (7.3) in that the superlative intensifies the negative evaluation (“complaint”, “drawback”) of the features *battery life* and *software*, however only the former received a feature label (battery life[-3]). Given the structural and semantic similarity of the examples, one could clearly argue for adding a feature label “software[-3]” to (7.6).

(7.4) compared to everything else in this category , {this} is most definitely [the] *best* [bang for the buck] .

(7.5) i did a good month ’s worth of research before buying this over other similar priced digital cameras , and {this} is [the] *best* [buy for the buck] .

(7.6) [the] *biggest* [drawback that people have about the zen xtra] is {the software} .

## 7.5 Experiment 4: Identifying ISA-1 superlatives in customer reviews

### 7.5.1 Overview

Having established that ISA superlatives are special indicators for product features in customer reviews, the aim of this experiment is to show that SUP-Classifier can be reliably used to identify ISA-1 superlatives in customer review data. Due to the variation in language, non-standard spellings (as well as typing errors), and the likely tagging/parsing errors that result from this, the initial expectation is that its performance will be considerably worse than on Wikipedia texts.

### 7.5.2 Results

SUP-Classifier is tested on the 230 superlative-containing sentences in Hu and Liu's corpus of customer reviews. Five of these sentences are excluded from evaluation as the C&C parser failed to parse them, leaving a total of 225 sentences for evaluation. The results are displayed in Table 7.4.

Class	Precision	Recall	F-Measure
<b>IDIOM</b>	<b>100.0%</b> (4/4)	<b>36.7%</b> (4/11)	<b>53.7%</b>
<b>PP</b>	<b>100%</b> (26/26)	<b>96.3%</b> (26/27)	<b>98.1%</b>
<b>PROP</b>	<b>89.8%</b> (44/49)	<b>95.7%</b> (44/46)	<b>92.7%</b>
<b>ADV</b>	<b>60.0%</b> (6/10)	<b>60.0%</b> (6/10)	<b>60.0%</b>
<b>FREE</b>	<b>33.3%</b> (1/3)	<b>33.3%</b> (1/3)	<b>33.3%</b>
<b>INDEF</b>	<b>63.2%</b> (12/19)	<b>80.0%</b> (12/15)	<b>70.6%</b>
<b>ISA-1</b>	<b>94.6%</b> (53/56)	<b>85.5%</b> (53/62)	<b>89.8%</b>
<b>ISA-2</b>	<b>33.3%</b> (1/3)	<b>12.5%</b> (1/8)	<b>18.2%</b>
<b>DEF</b>	<b>68.4%</b> (39/57)	<b>90.7%</b> (39/43)	<b>78.0%</b>
Average	71.4%	65.6%	66.0%

Table 7.4: Results of SUP-Classifier in Hu and Liu (2004)

With 94.6% precision and 85.5% recall for ISA-1 superlatives, the results show that SUP-Classifier can be reliably used to identify ISA-1 superlatives in customer re-

views. The non-standard nature of the data does not seem to have had a negative effect on `ISA1-Identifier`. In fact, its performance is better than on the `TextWiki` test set (82.4% precision and 84.3% recall). Before investigating these results in more detail, a short summary of the results of the other classes is given in the next paragraph.

Not all of the classes have better results on the customer review data. Compared with the results of `SUP-Classifier` on `TextWiki` (Table 5.21), its average performance on Hu and Liu’s corpus is around 10% lower, both in terms of precision (71.4% vs. 80.6%) and recall (65.6% vs. 75.5%). However, this does not apply to all classes equally: Apart from `ISA-1`, four other superlative classes have better F-measure values in the `Hu&Liu` test set than in `TextWiki`. While the performance of the `PP-` and `DEF-Identifiers` are a 2-3 percent points higher, `PROP-Identifier` achieves an F-measure improvement of around 5%. Most striking is the F-measure value of the `INDEF` class with 70.6%, which is 30.6% better than on the `TextWiki` data set. The lower average results are therefore due to the `IDIOM`, `ADV`, `FREE`, and `ISA-2` classes. These are all of relatively low frequency, and together account for only 14.2% of all superlatives (cf. Table 7.2). Therefore, when comparing the *overall* (rather than the *average*) precision and recall values (as shown in Table 7.5), `SUP-Identifier`’s results are not much worse on the customer review data.

Data set	Overall precision	Overall recall	Overall F-Measure
TextWiki	85.0% (334/393)	85.0% (334/393)	85.0%
Hu&Liu	81.9% (186/227)	82.7% (186/225)	82.3%

Table 7.5: Overall precision and recall

### 7.5.3 Discussion

This section discusses results for the `ISA-1` and `ISA-2` classes, which have been shown to be relevant for feature identification in customer reviews. A confusion matrix along with a discussion of the performance of all classes can be found in Appendix D.1.

The fact that `ISA1-Identifier` shows better performance on customer review data than on `TextWiki` may be surprising. However, a closer investigation of the gold-standard `ISA-1` superlatives (cf. column 2 in Table E.2) shows that while the system has to cope with non-standard language, the syntactic structure of `ISA-1` cases tends to

be less complex than in TextWiki (and WSJ). It is characterised by short and straightforward phrases of the form “X is the best Y”, and there are hardly any appositions in the set (which were shown to be a problem for SUP-Classifier in previous chapters).

Table 7.4 shows that nine out of 62 ISA-1 instances were not recognised. While most of the errors are due to similar reasons as the ones discussed in SUP-Classifier’s evaluation on Wikipedia data, some of them are a direct result of the C&C tool’s inability to handle the non-standard nature of customer review language. For example, (7.7) is classified as INDEF because the module `findDeterminer` fails to identify “it’s” as erroneous variant of the possessive pronoun “its”, incorrectly tagged as personal pronoun (PRP) + 3rd person singular present tense verb (VBZ). Example (7.8), on the other hand, is not recognised by the parser because of the phrase “just about...”, where “about” is interpreted as preposition rather than as a preceding adverb.

(7.7) i think this is it<sub>{PRP}</sub>’s<sub>{VBZ}</sub> *biggest* flaw .

(7.8) if you do any research into digital cameras , you ’ll quickly find that this camera is just<sub>{RB}</sub> about<sub>{IN}</sub> the *best* value out there .

The ISA-2 class is the one with the worst recall: Only one out of eight ISA-2 superlatives was recognised by the Identifier (12.5%). A closer investigation of the seven errors shows that they are the result of a variety of problems. As in Chapter 5.3.5, there are cases where an ISA-2 indicator is missing from its respective lexicon, like “seem” in (7.9). There are, however, also a number of cases where the error is due to additional elements inserted in the pattern, such as “like” in (7.10) and “among” in (7.11). While the latter can be solved by including a further rule in the set of “Minor methods” (as discussed in Chapter 5.3.5), the former needs to be addressed separately, possibly via a new ISA-2 pattern involving the word “like”.

(7.9) i looked into buying an inexpensive dvd player that had more than the standard set of features and this item seemed to be the *best* in that category .

(7.10) based on the cameras features and about dozen online reviews , this one seemed like the *best* all round deal .

(7.11) in fact , i 'd have to rate it among the *worst* i 've ever encountered .

Another group of ISA-2 errors results from tokenisation or tagging errors. As ISA2-Identifier strongly relies on the GR output, small errors on any of the underlying annotation levels can cause the Identifier to fail. Example (7.12), which should match ISA-2 Pattern I (“verb + T + CS”), is problematic not only because of the non-standard forms “ive” for “I’ve” and “em” for “them” (tagged as JJ and NN respectively), it should also have been tokenised as two separate sentences, as the GR parser incorrectly considers the verb “make” to be in a *dobj* relation with “everything”, and interprets “this” to be a determiner of “one”. This means that “one” is not recognised as CS head, and “this” is not found as target. In (7.13), on the other hand, the GR parser does not recognise “the” (underlined) to be the CS determiner. Instead, the CS head “bargain” is interpreted as direct object to “the” (the relevant GR tuples is (*dobj* the\_7 bargain\_9)).

(7.12) ive seen em all , the archos , dell , ipods , everything ... good price and great sound makes this one the *best* in my opinion .

(7.13) recent price drops have made the g3 the *best* bargain in digital cameras currently available .

Altogether, with a precision value of 94.6% and an F-measure of 89.8%, it can be concluded that SUP-Classifier can be reliably used to identify ISA-1 superlatives in customer reviews. The Identifier dealing with ISA-2 superlatives, on the other hand, needs some improvement, as already noted in earlier chapters (5.3.5).

## 7.6 Experiment 5: Identifying T and CS of ISA-1 superlatives in customer reviews

### 7.6.1 Overview

Having identified ISA-1 superlatives in the customer reviews, the next step is to extract their components so that potential product features can be detected. For this purpose, T/CS-Identifier is run on the set of gold standard ISA-1 superlatives, which have

been annotated with target and comparison set spans according to the guidelines in Chapter 4.2.2.

## 7.6.2 Results

Table 7.6 displays the results of running T/CS-Identifier on the customer review data set.

System	Accuracy
CS-Identifier	62.9% (39/62)
- CSDet	98.4% (61/62)
- CSMain	64.5% (40/62)
T-Identifier	66.1% (41/62)
T/CS-Identifier	45.2% (28/62)

Table 7.6: Performance of T/CS-Identifier and its components on ISA-1 superlatives in Hu&Liu (Accuracy)

Compared with T/CS-Identifier’s performance on TextWiki, one can see that both components perform worse on customer reviews. While T-Identifier’s accuracy is only around 3% lower, the results of CS-Identifier are almost 30% lower than on TextWiki. The component identifying the determinative phrase still has excellent results (98.4%), but the CSMain Identifier only correctly recognises 40 out of 62 spans. Interestingly, T-Identifier’s results are now better than those of CS-Identifier. This will be further discussed below. Overall, T/CS-Identifier manages to extract 28 correct sets of T and CS (45.2%), which is 17.% lower than its performance on TextWiki.

## 7.6.3 Discussion

The only CSDet span that was not recognised is shown in (7.14) and was already discussed in the evaluation of SUP-Classifier, as the erroneous spelling of “its” as “it’s” caused the Classifier to fail.

(7.14) i think {this} is [it ’s] *biggest* [flaw] .

The 22 errors in the CSMain span can broadly be divided up into three main groups: Errors caused by the tagger/parser (8), errors caused by the tagger/chunker (4), and errors caused by CS-Identifier (10). The eight errors of the tagger/parser are all of a similar kind, and exemplified by (7.15). In “bare” relative clauses starting with the pronoun “i” that are directly attached to the CSHead (e.g. “player” in 7.15), the parser falsely interprets “i” as the NP head, therefore causing T/CS-Identifier to label it as CSHead. This is due to the non-standard spelling of “i”, which caused it to be tagged as NNS (plural noun) instead of PRP (personal pronoun). (A quick test confirmed this: Running the same sentence through the tagger with “I” capitalised resulted in the correct analysis.)

(7.15) {this} is [the] *best* [dvd player i ’ve purchased] .

The chunking errors concern the same problem. Here, the CSHead has been identified correctly, but the personal pronoun “i” is taken to occur *inside* an NP chunk (I-NP), which causes the `relClause_BNP` method (which would otherwise be responsible for identifying the relative clause) to fail. This problem not only affects bare relative clauses starting in “i”, as sentence (7.17) illustrates. Here, the token “nokia” was not recognised as the start of a new NP chunk (B-NP), which usually indicates the start of a relative or subordinate clause.

(7.16) {this} is by far [the] *finest* [camera in its price and category i have ever used]

(7.17) {this} is [one of the] *nices* [phones nokia has made] .

The errors caused by CS-Identifier are due to incorrectly treated commas or “and” tokens (7.18), cases where the `addSuperlativeTag` method wrongly assigned the I-SUP tag to the token following “most” (7.19), and prepositions introducing subordinate clauses, which are not yet considered by the Identifier (7.20).

(7.18) in fact , my boyfriend is now going to invest into one of these suckers as well ... now he realizes that {this} is [the] *best* [choice over his current archos and the ipod he was eyeing] .

(7.19) {it} is [the] *most* [bang-for-the-buck out there].

(7.20) in my opinion {it} 's [the] *best* [camera for the money if you 're looking for something that 's easy to use] , small good for travel , and provides excellent , sharp images .

As noted above, despite the non-standard nature of the data, T-Identifier has better results on customer reviews than on TextWiki or Bos and Nissim's WSJ texts. A closer look at the data reveals two main reasons for this: Firstly, the target location methods work well, as sentences are generally short, and there are fewer appositions. The majority of targets were located via ISA1-Identifier's Method Ib, where the CS occurs in `xcomp` position (45 out of 62 cases). The second reason for the better results is that the target heads are often pronouns which have neither pre- nor post-modifiers. This applies to a total of 30 instances (almost half), with the demonstrative pronoun "this" occurring as THead 20 times, and the personal pronoun "it" occurring as THead 10 times. Furthermore, there are another 15 instances where THead consists of a simple NP such as "Apex" or "this dvd player". None of these represent a problem to T-Identifier (as long as the target has been correctly located).

The fact that a large proportion of targets are represented by pronouns immediately raises the question of pronoun resolution, and whether T-Identifier is of any use if the name of the actual target entities largely remain unknown. However, a first investigation of the data suggests that the great majority of the pronouns "this" and "it" refer to the entity under review. (This claim would however have to be verified by a thorough investigation of the context.) With respect to the goal of the current experiments (i.e identifying product features), pronouns in the target string do not represent a problem, as most product features occur in the comparison set string.

Having successfully identified ISA-1 superlatives and extracted their components, the important final question is how these results can be used to arrive at the product features they are assumed to contain. As mentioned above, the feature is a substring of either the target or the comparison set in 34 out of the 46 instances (73.9%). As the majority of them (27) occur as part of the comparison set, one strategy would be to assume that the product feature substring is the NP-chunk containing the CSHead. This simple approach would work for 25 of the 27 cases. Crucially, as most of the errors

in automatically detecting the CS span were caused by the postmodifier modules of CS-Identifier, the system will still be able to correctly identify product features, as only the CSHead (plus surrounding NP-chunk) is required (but no information about the postmodifiers).

## 7.7 Conclusion

This chapter described a first application of the findings of this PhD in the area of Sentiment Analysis/Opinion Extraction. It found ISA-1 superlatives to be special indicators of product features, which not only contain the feature strings (in most cases as part of the CS), but also the opinion word (usually the superlative itself)<sup>3</sup>, thus solving two Opinion Mining tasks at once. As this strategy for finding product features does not depend on their frequency (unlike Hu&Liu’s approach), it also represents an efficient way of locating the so-called *infrequent* features, which are also of great interest in Opinion Mining. The two experiments described in (7.5) and (7.6) show how the software developed in this PhD project can be used to reliably identify ISA-1 superlatives, and to extract from them potential product feature strings.

While this chapter has focused on the role of ISA-1 superlatives in Opinion Mining, it is worth pointing out that another interesting and potentially useful class is represented by DEF, as illustrated by (7.21) and (7.22), which express positive statements about the features “image quality” and “lens adapter”, respectively.

(7.21) overall , the g3 delivers what must be considered the *best* [image quality of any current > 4 megapixel digicams , from a detail , tonal balance and color response point of view] .

(7.22) they got the *best* [lens adapter] for the g3-better than canon ’s .

While the distribution of product features across the DEF class does not hint at their importance (cf. Table 7.3), one needs to bear in mind that the DEF class contains a wide variety different semantic types (as discussed in Chapter 3.2.2.2), of which only

<sup>3</sup>A special case is represented by superlatives that modify nouns such as “flaw” or “advantage” (e.g. 7.14), which are opinion words themselves. Such superlatives do not express properties shared by the members of the set, but have the role of intensifying the positive or negative meaning of the nouns (cf. Chapter 1.2.2).

the so-called “relative set comparisons” type may be of interest. Future work may therefore involve finding techniques that can distinguish between this type of DEF superlative and the other semantic types found in this class.

# Chapter 8

## Conclusion

This chapter summarises the main findings and contributions of this thesis and outlines future research directions.

### 8.1 Summary and main findings

This thesis proposed a computational treatment of superlatives, dealing with the main challenges in automatically recognising and extracting their components. The following is a summary of the central findings and contributions of this work.

Superlatives are of interest from a computational perspective because they express a comparison between a target entity and its comparison set. For example, in the sentence “The blue whale is the largest mammal”, the target *blue whale* is compared to the comparison set of *mammals*. My initial investigation of superlative forms showed that two types of relation hold between a target and its comparison set: An *IS-A relation*, which expresses the membership of the target in the comparison class, and a *superlative relation*, which specifies a property that all members of the set share, but which the target has the highest (or lowest) degree or value of. Both of these relations are of great interest from a relation extraction point of view, and the question this thesis aims to address is that of automatically extracting useful information about the target entity, its comparison set and their relationship from superlative constructions. In Chapter 1, the task was formulated to consist of the following three subtasks:

- **TASK 1:** Decide whether a given sentence contains a superlative form

- **TASK 2:** Given a sentence containing a superlative form, identify what type of superlative it is
- **TASK 3:** For set comparisons, identify the target and the comparison set, as well as the superlative relation

An investigation of the types of comparisons superlatives can express showed that it is not possible to identify a unified computational treatment for all types due to their great semantic complexity and the variety of surface structures in which they occur. For a computational treatment of superlatives that has as its goal to extract the main components of a superlative comparison (i.e. information about the target and comparison set), it is of vital importance to distinguish between the different types of comparison that superlatives can express, and to identify the ones whose components lend themselves most readily for extraction. For this reason, I proposed a classification of superlative surface forms, and decided to focus on so-called “ISA superlatives”, which make explicit the IS-A relation that holds between target and comparison set. They are suitable for a computational approach because both their target and comparison set are usually explicitly realised in the text.

In order to deal with the three tasks above automatically, a corpus was required where superlatives are annotated with their class membership, and where the target and comparison set strings of all instances classified as ISA are marked up. For this purpose, I devised an annotation scheme for these two tasks, which was tested and evaluated on 500 tokens of superlatives. In addition to providing a platform for investigating superlatives on a larger scale, this also introduced a new text-based Wikipedia corpus (“TextWiki”) which is especially suitable for linguistic research.

Three software components were implemented in the framework of this project which offer an efficient way of dealing with superlatives automatically. The “Superlative Relation Extractor” (SRE) contains a tool which can identify superlatives in free text with high precision and recall, addressing TASK 1 above (SUP-Finder). The second component is able to reliably classify superlative instances according to the surface form classification mentioned in the previous paragraph, solving TASK 2 (SUP-Classifier). One of the classes it can recognise is ISA-1, whose members (as noted above) lend themselves most readily for an extraction of their components. For any member of this class, the third component that was developed as part of the project is able to identify the target and comparison set span, thus solving TASK 3 (T/CS-Identifier). All

components integrated in SRE were developed, tested, and evaluated on the `TextWiki` corpus with very good results.

Finally, the work described in this thesis has been shown to be useful for a variety of NLP areas. I discussed possible applications within the areas of Question Answering, Ontology Learning, and Natural Language Generation, showing ways in which they could benefit from an automatic treatment of superlatives. The last chapter of the thesis described two experiments which demonstrate a first practical application of this work in the area of Sentiment Analysis/Opinion Mining. It established that `ISA-1` superlatives are special indicators of product features, the identification of which is a major task in Opinion Mining. The chapter showed that `SUP-Classifier` and `T/CS-Identifier` can be used with good results to find product features in customer reviews.

## 8.2 Future research directions

In the following, I outline future work which addresses the limitations of this work as well as the further development of the ideas described in this thesis.

The exhaustive **classification of superlative surface forms** proposed in this thesis provides a useful platform for a closer theoretical investigation of the semantics of superlative comparisons. While Chapter 3 provided an overview of some main semantic types of comparison, these yet have to be investigated in more detail. From the point of view of theoretical linguistics, the `TextWiki` corpus represents an interesting resource for the study of the semantics of superlatives, as all superlative instances have been identified and annotated with class information. From a computational point of view, an investigation of the `DEF` class would be particularly interesting, as it contains relative set comparisons, which were shown to be of great potential interest in areas such as Opinion Mining. Future work might also see the implementation of a further classifier which can distinguish relative set comparisons from the other semantic types in the `DEF` class.

The automated **identification of the target and comparison set** of superlatives also offers interesting future research directions. First of all, the existing Identifier can be improved by addressing the problems identified in its evaluation on the Wikipedia, WSJ, and customer review data. Secondly, further annotations and methods could be

developed that address the following issues:

1. *Semantic role labelling*. In order to interpret the dependants of the superlative comparison semantically, a suitable classification of semantic roles is required. The constituents identified by `T/CS-Identifier` could serve as a useful basis for this (Chapter 6). In particular, one major difficulty will lie in determining the difference between NP-head complements and NP-head modifiers, defined by Huddleston and Pullum (2002) as follows:

[PRE-HEAD]	
i. a linguistics student	(complement)
ii. a first-year student	(modifier)
[POST-HEAD]	
i. a report on the crash	(complement)
ii. a report in the paper	(modifier)

Here, cases [i] are complements of the NP head, while cases [ii] are modifiers. My first impression is that complements have the role of “defining” the members of the set (i.e. defining the entities “as such”), while modifiers have the role of restricting the set by only considering a subset of these members, for example those in specific locations or at specific times. A correct handling of these issues is crucial for the task. A first place to start will be two papers: One by Merlo and Leybold (2001), “Automatic Distinction of Arguments and Modifiers: the Case of Prepositional Phrases”, and the other by Merlo and Ferrer (2006), “The Notion of Argument in PP Attachment”.

2. *Anaphora resolution and preposed CS elements*. A great proportion of target and comparison sets require anaphora resolution. This issue seemed particularly prominent in the domain of customer reviews, where many targets were represented by pronouns. A further problem is represented by constituents that modify the comparison set but occur in a preposed position (e.g. prepositional phrases at the beginning of a sentence). Both problems need to be addressed to achieve useful semantic representations of target and comparison set.

3. *External restrictions*. In addition to preposed elements, ISA comparisons can also be restricted or modified via other external constituents, such as negation or adverbial phrases. These have been marked up in the `TextWiki` corpus and can be used to develop methods for their identification.

4. *Tenses*. So far, the treatment of ISA superlatives does not take different tenses of the

verb “to be” (or other copula verbs) into account. These have an effect on the reliability of the comparison: Past tense forms, for example, can imply that the opinions or facts expressed by the superlative comparison are no longer correct. In addition, a future version of the Identifier should also take the modality of verbs into consideration, as it may also affect the reliability of facts.

Finally, Chapter 7 indicated that superlatives play a special role in **Sentiment Analysis and Opinion Mining**, as they are used to express opinions about particular product features. However, the study did not consider whether the positive or negative opinion was based on a subjective impression or on a fact. An initial investigation suggests that the subjectivity status of superlatives is often borderline between subjective and objective. This is reflected by the fact that their use in advertising is often hotly debated and constitutes a current legal issue which is regulated via the “Advertising Standards Authority” (ASA). In advertising, superlatives are known to express vague or subjective opinions that exaggerate the function or effectiveness of products. The reason for this problem might be found on a pragmatic level: As superlatives rank entities according to the degree to which they possess a certain property, they are inherently factual. A conflict arises when the superlative comparison is based on a subjective evaluation: While the surface structure suggests the presentation of a fact, the semantics of the underlying adjective are in fact based on an opinion. This deceives customers, as by using a superlative comparison the customer feels that he/she is presented with a fact, when in reality the information is subjective. A pragmatic investigation of the use of superlatives in advertising and in customer reviews promises to be an interesting topic for future research, which will further benefit applications in Sentiment Analysis and Opinion Extraction.

# Appendix A

## Appendix to Chapter 2

### A.1 Superlatives in TREC

#### A.1.1 TREC 2002

- < 1397 > What was the *largest* crowd to ever come see Michael Jordan?
- < 1424 > Who won the Oscar for *best* actor in 1970?
- < 1433 > What is the height of the *tallest* redwood?
- < 1441 > What is the *coldest* place on earth?
- < 1462 > Where is the *oldest* synagogue in the United States?
- < 1474 > What is the *lowest* point on earth?
- < 1483 > Where is the *highest* point on earth?
- < 1503 > What is the world's *second largest* island?
- < 1529 > What is the *oldest* college bowl game?
- < 1540 > What is the *deepest* lake in America?
- < 1576 > What galaxy is *closest* to the milky way?
- < 1577 > What Burns poem does "the *best* laid plans of mice and men" come from?
- < 1602 > Who holds the record as the *highest* paid child performer?
- < 1615 > What is Africa's *largest* country?
- < 1625 > What is the *deepest* lake in the world?
- < 1633 > What roller coaster is the *fastest* in the world?
- < 1668 > What is the *oldest* national park in the U.S.?
- < 1743 > Which state has the *longest* coastline on the Atlantic Ocean?

- < 1770 > What is the *highest* recorded temperature in San Antonio, TX?
- < 1779 > What is Australia's *oldest* city?
- < 1793 > What is the world's *tallest* office building?
- < 1794 > What is the *fastest* car in the world?
- < 1801 > Where is the *smallest* bone in the body?
- < 1826 > Which film received the first *best* picture Academy Award?
- < 1846 > What is the *oldest* sports trophy?
- < 1879 > By what nickname was musician Ernesto Antonio Puente, Jr. *best* known?
- < 1397 > What was the *largest* crowd to ever come see Michael Jordan?
- < 1514 > What is Canada's *most populous* city?
- < 1544 > What is the *most populated* country in the world?
- < 1550 > What is the *southwestern-most* tip of England?
- < 1579 > Which country exports the *most* tea?
- < 1685 > What is the *most populous* city in the United States?
- < 1773 > What was Aaron Copland's *most famous* piece of music?
- < 1780 > Who has the *most* no hitters in major league baseball?
- < 1786 > What were the *most* points Michael Jordan scored in a game?
- < 1843 > In what month are the *most* babies born?

### A.1.2 TREC 2003

- < 1913 : *list* > What breeds of dog have won the "Best in Show" award at the Westminster Dog Show?
- < 1956 : *factoid* > What country is the *largest* in square miles?
- < 1982 : *factoid* > What movie won the Academy Award for *best* picture in 1989?
- < 1999 : *factoid* > What is the state with the *smallest* population?
- < 2005 : *factoid* > What is the *fastest* roller coaster?
- < 2021 : *factoid* > What is the world's *largest* coral reef?
- < 2023 : *factoid* > What continent is the world's *largest* dessert on?
- < 2028 : *factoid* > What are the *biggest* snakes in the world?
- < 2053 : *factoid* > What duo lost their Grammy for *best* new artist when it was discovered that they lip-synched their songs?
- < 2074 : *factoid* > How tall is the *tallest* pyramid?
- < 2081 : *factoid* > Which is the *longest* track event in the Olympic Games?

- < 2087 : *factoid* > What Canadian city has the *largest* population?
- < 2189 : *factoid* > What animal can go the longest without water?
- < 2230 : *factoid* > What dam is said to be the *largest* hydroelectric station in the world?
- < 2235 : *factoid* > What is the *highest* mountain in South America?
- < 2237 : *factoid* > What album became the *greatest* selling album of all time?
- < 2246 : *factoid* > What film did Liza Minnelli win a *best* actress Oscar for?
- < 2257 : *factoid* > What is the *richest* country in the world?
- < 2273 : *factoid* > Which religion has the *largest* number of followers worldwide?
- < 2303 : *factoid* > What is the *longest* river in the world?
- < 2316 : *factoid* > What is the *largest* city in Austria?
- < 2357 : *factoid* > What city's *biggest* shopping district is called the Ginza District?
- < 2374 : *factoid* > What is the *tallest* statue in the world?
- < 2068 : *factoid* > What animal is responsible for the *most* human deaths worldwide?
- < 2105 : *factoid* > What country produces the *most* silk?
- < 2115 : *factoid* > In what Canadian province are *most* people of French descent?
- < 2157 : *factoid* > What state had the *most* signers of the Declaration of Independence?
- < 2168 : *factoid* > What country produces the *most* emeralds?
- < 2331 : *factoid* > What actress has received the *most* Oscar nominations?

## A.2 “Other” question nuggets

### A.2.1 Superlative nuggets in TREC 2004/2005

Table A.1: Superlative nuggets in TREC 2004/2005

N	Qid	target	nugget	V	OK	type
1	3.4.2	Hale Bopp comet	one of the brightest comets of 20th century	10	0	S1
2	5.6.3	AARP	Largest seniors organization	10	0	S1
3	12.5.1	Rohm and Haas	worlds second largest specialty chemical manufacturer	10	0	S1
4	19.6.5	Kibbutz	Kibbutzim in Israel is largest communitarian movement in world today	10	0	S1
5	29.4.2	Tale of Genji	Japans greatest literary achievement	10	0	S1
6	41.5.1	Teapot Dome scandal	largest scandal in U.S. history	10	0	S1
7	45.4.2	International Finance Corporation (IFC)	largest multilateral source of loan financing for private sector projects in developing countries	10	0	S1
8	52.6.1	Burger King	2nd largest burger chain	10	0	S1
9	75.8.1	Merck & Co.	Worlds largest drug company	10	0	S1
10	77.8.5	George Foreman	Became oldest world champion in boxing history	10	0	S1

*Continued on Next Page...*

11	88.7.4	United Parcel Service (UPS)	worlds largest transportation firm	10	0	S1
12	98.6.3	American Legion	nations largest veteran organization	10	0	S1
13	102.7.3	Boston Big Dig	most complex public works project ever undertaken in history of U.S.	10	0	S1
14	104.8.3	1999 North American International Auto Show	Longest running auto show in the country	10	0	S1
15	109.6.3	Telefonica of Spain	Largest supplier of telecommunications services in Spanish speaking world	10	0	S1
16	110.7.3	Lions Club International	Worlds largest volunteer charitable organization	10	0	S1
17	111.6.3	AMWAY	One of the largest direct sales companies	10	0	S1
18	112.7.4	McDonalds Corporation	Largest Fast Food Company in United States	10	0	S1
19	115.8.1	Longwood Gardens	one of the two best public gardens in America	10	0	S1
20	118.7.1	U.S. Medal of Honor	highest military award for gallantry in action	10	0	S1
21	48.6.2	Abu Nidal	labelled worlds most dangerous terrorist	9	1	S1
22	67.7.2	Miss Universe 2000 crowned	Pageant 3rd most popular TV event world-wide	9	1	S2

*Continued on Next Page...*

23	68.8.2	Port Arthur Massacre	Shooting considered Australias worst mass murder	9	1	S1
24	68.8.3	Port Arthur Massacre	Was worlds worst massacre by a lone gunman	9	1	S1
25	78.8.4	Akira Kurosawa	Japans most famous film director	9	1	S1
26	99.7.2	Woody Guthrie	his most famous song, This Land is Your Land	9	1	S2
27	114.7.5	Jesse Ventura	Highest Elected Official in the Reform Party	9	1	S1
28	115.8.8	Longwood Gardens	one of the worlds largest conservatories	9	1	S1
29	116.7.1	Camp David	best known for peace agreement signed there between Egypt and Israel	9	1	S1
30	130.7.2	tsunami	Tsunamis most common in Pacific Ocean	9	1	S1
31	131.8.3	Hindenburg disaster	MOST FAMOUS AIR DISASTER	9	1	S1
32	133.6.1	Hurricane Mitch	The fourth-strongest Caribbean hurricane in this century	9	1	S1
33	15.4.10	Rat Pack	Rat Pack most famous entertaining troupe in history	8	2	S1
34	27.5.7	Jennifer Capriati	Youngest semi-finalist ever in Grand Slam tournament.	8	2	S1
35	33.5.1	Florence Nightingale	Nightingale Medal highest international nurses award	8	2	S2
36	35.5.4	Jack Welch	most admired businessman	8	2	S1
37	50.5.13	Cassini space probe	Cassini, NASAS Biggest and most complex interplanetary probe	8	2	S1
38	51.4.12	Kurds	worlds biggest stateless nation	8	2	S2
39	78.8.5	Akira Kurosawa	Best known film (translated) was THE SEVEN SAMURAI	8	2	S2

*Continued on Next Page...*

40	113.7.6	Paul Newman	At Age 70, he entered the Guinness [sic] Book of World Records as oldest to win a competitive sanctioned racing event	8	2	S1
41	5.6.4	AARP	Largest dues paying organization	7	3	S1
42	27.5.4	Jennifer Capriati	Youngest to reach earnings of \$1M.	7	3	S1
43	27.5.6	Jennifer Capriati	Youngest to have Top Ten ranking.	7	3	S1
44	31.9.2	Jean Harlow	Saratoga was her most successful film	7	3	S2
45	88.7.1	United Parcel Service (UPS)	largest IPO in stock market history	7	3	S1
46	100.8.5	Sammy Sosa	Sosa received his nations highest honor	7	3	S3
47	131.8.2	Hindenburg disaster	LARGEST AIRSHIP 100 TONS	7	3	S1
48	132.6.7	Kim Jong II	He is one of the worlds most reclusive leaders	7	3	S1
49	92.7.9	Arnold Palmer	most consecutive masters 44	7	3	S1
50	50.5.4	Cassini space probe	send Huygens to probe atmosphere of Titan, Saturns largest moon	6	4	S3
51	85.7.9	Norwegian Cruise Lines (NCL)	Is the most sports minded of cruise lines	6	4	S1
52	97.7.7	Counting Crows	crowns new album, this desert life, is among the best pop-rock efforts of the year	6	4	S2
53	124.7.11	Rocky Marciano	Among best of heavyweights	6	4	S1

*Continued on Next Page...*

54	136.8.6	Shiite	Imam Hussein is Shiite Islams most revered saint	6	4	S2
55	51.4.14	Kurds	Irbil largest city controlled by Kurds	6	3	S3
56	40.6.6	Chester Nimitz	Nimitz class carrier largest in fleet	5	5	S2
57	103.7.1	Super Bowl XXXIV	best super bowl ever	5	5	S1
58	105.7.3	1980 Mount St. Helens eruption	the most climbed volcano in North America	5	5	S2
59	130.7.6	tsunami	Eruption Krakaton 1883 caused worst Tsunami	5	5	S2
60	95.6.3	return of Hong Kong to Chinese sovereignty	Largest Trade Partners	4	6	S3
61	130.7.9	tsunami	Tsunamis deadliest threat in Hawaii	4	6	S1
62	100.8.12	Sammy Sosa	most memorable home run chase ever	4	5	S2
63	130.7.11	tsunami	Japan most affected by tsunami 250 100,000 deaths	4	5	S3
64	98.6.8	American Legion	Schick, oldest known living member of the Legion	3	7	S2
65	111.6.8	AMWAY	\$120 million plant in Guangzhou China is their largest overseas production center	3	7	S2
66	128.6.7	OPEC	Iran second largest oil producer	3	7	S3
67	79.7.2	Kip Kinkel school shooting	named most likely to start World War III	3	7	S2
68	98.6.11	American Legion	legion says areas with greatest need for new national cemeteries include okla. city, Atlanta, sacramento	0	10	S3

Continued on Next Page...

69	103.7.11	Super Bowl XXXIV	Mike Horan, at 39 one of the oldest players in Super Bowl history	0	10	S3
----	----------	------------------	---	---	----	----

### A.2.2 Comparative nuggets in TREC 2004/2005

Table A.2: Comparative nuggets in TREC 2004/2005

<b>N</b>	<b>Qid</b>	<b>target</b>	<b>nugget</b>	<b>V</b>	<b>OK</b>
1	128.6.1	OPEC	OPEC founded in Iraq 1960 for higher oil prices	9	1
2	68.8.1	Port Arthur Massacre	Massacre triggered consideration of tighter gun control	5	5
3	99.7.13	Woody Guthrie	His voice [sic] was once an untamed cockney bray, has grown more controlled and gentle	2	8
4	39.4.4	The Clash	Rancid plays faster than Clash	1	9
5	68.8.6	Port Arthur Massacre	Australia considered safer after massacre inspired gun control	1	9

# Appendix B

## Appendix to Chapter 3

### B.1 Definitions

#### DEFINITION D.1 (*idiomatic superlative*):

An idiomatic superlative is a superlative form that is in some way lexicalised. Idiomatic superlative forms are usually not produced by an act of “active” grading (i.e. by an “on-the-fly” derivation of a superlative form from the base adjective/adverb), but are part of a fixed phrase or expression. We will distinguish between the following classes of idiomatic superlatives:

#### 1. Proportional quantifiers

- phrases of the form **most** + (*of* + (DET)) + (NP)
- e.g. **Most of the books** (*are boring*); **Most children** (*hate spinach*).

#### 2. Prepositional Phrase superlatives

- phrases of the form [preposition + (*the*) + (*very/...*) + superlative]
- e.g. *at most*, *at least*, *at the very least*, *at best*, *at worst*, *at the earliest*, *in the slightest*, *to its fullest*, ...

#### 3. Other idiomatic superlatives

- other superlative forms with idiomatic usage
- e.g. *last but not least*, *the Best Western Hotel*, *I will do my best*.

**DEFINITION D.2 (*adjectival vs. adverbial superlative*):**

Adjectival superlatives are superlatives that are:

- derived from adjectives as described in Chapter 1.3.1
- listed in Table 1.1 in Chapter 1.3.1

Adverbial superlatives are superlatives that are:

- derived from adverbs as described in Chapter 1.3.1
- listed in Table 1.2 in Chapter 1.3.1

**DEFINITION D.3 (*NP-bound superlative*):**

A superlative form is incorporated in an NP (and referred to as *NP-bound superlative* or *superlative NP*) if:

1. The superlative form occurs before the NP head in premodifying position, for example:
  - One of the *most common* gazelles in northern Africa [...]
  - The *largest* gibbon [...]
2. The superlative form is part of a so-called “fused head” construction, where the superlative and the NP head “merge” into one unit (Huddleston and Pullum 2002). The NP head is implied in the context (usually, but not necessarily in the same sentence). For example:
  - The Airedale is the *largest* of the terrier class.
  - The *largest* is the black caiman of the Orinoco and Amazon rivers.

**DEFINITION D.4 (*definite/indefinite superlative NP*):**

A *definite* superlative NP has one of the following as determiner:

- a definite or demonstrative determiner (*the/this/that/these/those*)
- a possessive pronoun or NP (*my/your/his/her/its/their/whose, Peter’s, the country’s*, etc.)
- a relative or interrogative determiner (*which*).

Otherwise, the superlative NP is called *indefinite*.

# Appendix C

## Appendix to Chapter 5

### C.1 ISA1-Identifier: Analysis of false negatives

All 14 instances that were not recognised by ISA1-Identifier were classified as DEF.

- (C.1) The centre has been praised for its architecture and state-of the-art modern theatre facilities, among the *best* in the world, but criticized for its bureaucratic arts administration.

[“among CS” recogniser does not work in combination with apposition]

[*New National Theatre*]

- (C.2) The *largest* and most active year-round youth theater in Connecticut, the Playhouse is an independent, non-profit organization with an arts/social service mission.

[*Module findCoord succeeds; but this type of ncsbj apposition is not recognised*]

[*Oddfellows Playhouse*]

- (C.3) The largest and *most* active year-round youth theater in Connecticut, the Playhouse is an independent, non-profit organization with an arts/social service mission.

[*This type of ncsbj apposition is not recognised*]

[*Oddfellows Playhouse*]

- (C.4) The *highest* point in the archipelago, Conachair ('the beacon') at 430 metres (1,411 ft), is on Hirta, immediately north of the village.

[“Conachair” is recognised as potential apposition to the CS head “point” (Method IIa, Step 2), but finding the shared verb “is” fails, because (like the previous two examples) this type of *ncsubj* apposition is not recognised]

[St Kilda, Scotland]

- (C.5) Surgery costs vary from country to country, with the US typically being among the *highest* markets, and countries like Thailand, Cuba and Argentina, among the lowest.

[“among CS” construction is detected (Method IV), but Identifier fails to recognise the target “US”]

[Hip replacement]

- (C.6) Surgery costs vary from country to country, with the US typically being among the highest markets, and countries like Thailand, Cuba and Argentina, among the *lowest*.

[“among CS” construction is detected (Method IV), but Identifier fails to detect coordination]

[Hip replacement]

- (C.7) The Human Genome Project - the *largest*, most costly single biological study ever undertaken - began in 1988 under the leadership of James D. Watson, after preliminary work with genetically simpler model organisms such as *E. coli*, *S. cerevisiae* and *C. elegans*.

[ISA1-Identifier (and the parser in this case) does not recognise apposition involving dashes]

[History of biology]

- (C.8) The Human Genome Project - the largest, *most* costly single biological study ever undertaken - began in 1988 under the leadership of James D. Watson, after preliminary work with genetically simpler model organisms such as *E. coli*, *S. cerevisiae* and *C. elegans*.

[*ISA1-Identifier (and the parser in this case) does not recognise apposition involving dashes*]

[*History of biology*]

- (C.9) However, it was only from 1735, with the arrival of the *most* illustrious of French governor, Mah de La Bourdonnais, that the “Isle de France” started developing effectively.

[*Apposition not recognised as it is not in ncsbj or dobj position; in addition, findHead identified “illustrious” as fused head, but parser treats “govenor” as part of conjunction*]

[*History of Mauritius*]

- (C.10) Upon his death, Qutb-ud-din Aybak, Muhammad Ghori’s *most* capable general, who had started of by sacking Ayodhya in 1193 A.D., took control of Muhammad’s Indian conquests and declared himself the first Sultan of Delhi thus establishing Sultanate of Delhi in 1206 CE.

[*Apposition not recognised by parser: no conj relation linking “Aybak” and “general”*]

[*Muhammad of Ghor*]

- (C.11) After his death, Kalpaks was posthumously awarded Latvia’s *highest* military award, ‘Order of Lacplešis I III Class’.

[*Apposition not recognised by parser: no conj relations in GR output*]

[*Oskars Kalpaks*]

- (C.12) When Baldwin IX, Count of Flanders and Hainault, left on the Fourth Crusade in 1202, he left his western domains under his *eldest* daughter Joanna.

[*This type of apposition is not supported by ISA1-Identifier*]

[*War of the Succession of Flanders and Hainault*]

- (C.13) The *largest* animal phylum belongs here, the Arthropoda, including insects, spiders, crabs, and their kin.

[*Apposition not recognised by parser*]

[Animal]

- (C.14) The Lophotrochozoa include two of the *most* successful animal phyla, the Mollusca and Annelida.

[*Apposition not recognised by ISA1-Identifier because findHead failed to recognise “two” as CS head*]

[Animal]

## C.2 ISA2-Identifier: Analysis of false negatives

List of ISA-2 instances that were not recognised by the Identifier because their (verbal) indicator was missing from the respective lexicon.

- (C.15) Tzara, in 1918, wrote a Dada manifesto considered one of the *most* important of the Dada writings.

[*consider* missing from Lexicon VIII]

[Dada]

- (C.16) The committee presiding over Britain’s prestigious Turner Prize in 2004, for example, called it “the *most* influential work of modern art.”

[*call* missing from Lexicon I]

[Dada]

- (C.17) RPGFan called Final Fantasy Adventure one of the *best* things to happen to the Game Boy.

[*call* missing from Lexicon I]

[*Mana (series)*]

- (C.18) GameSpy called Children of Mana’s music some of the *best* Nintendo DS music yet and referred to it as “beautiful”.

[*call* missing from Lexicon I]

[*Mana (series)*]

- (C.19) For example, one of the *most* famous early maps of North America is unofficially known as the Beaver Map, published in 1715 by Herman Moll.

[*know* missing from Lexicon VII]

[*Cartography*]

- (C.20) The *simplest* technique, of adding up the pixel counts within a circle centered on the object and subtracting off an average sky count, is known as aperture photometry.

[*know* missing from Lexicon VII]

[*Photometry (astronomy)*]

- (C.21) Analysis using electromyography has shown the bicycle crunch to be one of the *best* forms of abdominal muscle exercise.

[*show* missing from Lexicon IV]

[*Bicycle crunch*]

- (C.22) John Newton visited the area in 1743 and remarked that they seemed "the *most* humane and moral people I ever met with in Africa; and they were the people who had the least intercourse with Europe at that time".

[*seem* missing from Lexicon II]

[*Kingdom of Orungu*]

### C.3 DEF-Identifier: Analysis of false negatives

Analysis of cases falsely classified as ISA-1.

Clear errors:

- (C.23) I always rush up and take off ninety miles an hour for the *nearest* whorehouse, hor-hor-hor!

[*ISA1-Identifier falsely recognised "hor-hor-hor" as apposition.*]

[*Orgone*]

- (C.24) The *most* serious symptoms of altitude sickness are due to edema (fluid accumulation in the tissues of the body).

[“*due to edema*” is not a target.]

[*Altitude sickness*]

- (C.25) The *best* rules of the road allow any slower driver (including a cyclist) to establish the center of the outermost marked lane (between the left and right tracks of wider vehicles) as their default or primary position.

[*ISA1-Identifier falsely recognised “established” as target (parsing error).*]

[*Vehicular cycling*]

- (C.26) Often, at the end of the regular season, the league holds a post-season tournament (most commonly called a playoff) to determine which team is the *best* out of all of the other teams in the league.

[*ISA1-Identifier determined “which” as target. A new rule should be added that excludes WH-determiners from consideration for targets.*]

[*Bracket (tournament)*]

- (C.27) The *earliest* recorded attempts at hip replacement (Gluck T, 1891), which were carried out in Germany, used ivory to replace the femoral head (the ball on the femur).

[*ISA1-Identifier falsely recognised “replace” as target (parsing error).*]

[*Hip replacement*]

Borderline cases:

- (C.28) The *shortest* distance; that is, a straight channel, results in the highest energy per unit of length, disrupting the banks more, creating more sediment and aggrading the stream.

[*ISA1-Identifier found “that” as target. Indeed this could be an indicator of a following target, but gold standard annotation decided that the phrase “a straight channel” does not count as proper target.*]

[Meander]

(C.29) The *oldest* structures on St Kilda are the most enigmatic.

[ISA1-Identifier found “enigmatic” as target. From a semantic point of view, however, this does not count as a proper target.]

[St Kilda, Scotland]

(C.30) The oldest structures on St Kilda are the *most* enigmatic.

[ISA1-Identifier found “structures” as target. As in the previous example, this does not count as a proper target for semantic reasons.]

[St Kilda, Scotland]

(C.31) The *most* common are:

[Another borderline case. Strictly seen, it is an ISA-1 superlative as the target must be mentioned in the context. However, the target is not mentioned in the same sentence, which is why gold standard annotation labelled the example as DEF.]

[Bracket (tournament)]

(C.32) When the president steps down, by convention, the *most* senior member who has not been president yet, who, also by convention, is the vice-president, is elected president of the STF.

[Borderline case because a target is mentioned (“the vice-president”).]

[Supreme Federal Court (Brazil)]

# Appendix D

## Appendix to Chapter 6

### D.1 Description of PostMod rules

This section discusses in detail the PostMod rules shown in Table 6.11.

Module I, **ppPhrase**, tackles prepositional phrases, and is triggered when the focus token carries the POS tag IN (for “preposition” or “subordinating participle conjunction”), or the chunk tag B-PP or I-PP. Using chunk tags alone to identify prepositional phrases is of no more use than using POS tags, as the chunker has been trained to focus on noun phrases (NPs). This means that in an example like “in the house”, the chunker usually only labels the preposition as I-PP (or, rarely, B-PP), but treats the rest of the PP as an NP chunk: “*in/I-PP the/I-NP UK/I-NP*”.

Once a token has been recognised as preposition, a specially created lexicon is consulted which lists **ppPhrase** exclusions and restrictions. Excluded are all instances of “although”, “as”, “because”, “despite”, “except”, and “that”, because they are unlikely to describe a PP that restricts the comparison set.<sup>1</sup> A number of other prepositions are listed in the lexicon with restrictions: For example, “at” is excluded from consideration if it is followed by a cardinal number (CD), as it is then likely to be non-restrictive. For example, in (D.1), the preposition “at” is followed by “53.9 km”, which specifies the length of the tunnel, but does not restrict the comparison set. Similarly, “with”-PPs are only included if the second token of the phrase is not tagged IN and is not an indefinite determiner (*a/an*), which (correctly)

---

<sup>1</sup>The token “that” is treated separately in the `relClause` module.

excludes “with about one-fifth world share” in (D.2), but includes “with only the right hand” in (D.3). (In all examples, the focus token is highlighted in bold font).

- (D.1) The Seikan Tunnel in Japan is the *longest* [rail tunnel<sub>{CSHead}</sub> in the world] **at**<sub>{IN}</sub> 53.9<sub>{CD}</sub> km (33.4 miles), of which 23.3 km (14.5 miles) is under the sea.

[Tunnel]

- (D.2) In 2005, Russia was the *largest* [producer<sub>{CSHead}</sub> of nickel] **with**<sub>{IN}</sub> about<sub>{IN}</sub> one-fifth world share closely followed by Canada, Australia and Indonesia, as reported by the British Geological Survey.

[Nickel]

- (D.3) The word “hypolimnion” is one of the *longest* [words<sub>{CSHead}</sub> that can be typed **with**<sub>{IN}</sub> only<sub>{RB}</sub> the right hand on a QWERTY and AZERTY keyboard].

[Hypolimnion]

A further example is the word “since”, which can not only be used to denote periods of time (“since 1977”, “since the first day of the year”), but it can also be used in a similar way to “as” and “because” to provide a reason for a situation or action. In the latter case it is usually unrestrictive. As “since” can be followed by an NP in both cases, finding a method for distinguishing them is not straightforward.

CS-Identifier resorts to excluding cases of “since” that are followed by a personal pronoun (tag PRP). These are likely to refer back to the CSHead, giving a reason for or explanation of the superlative statement (as in D.4). All other cases of “since” are treated as potential PP-phrase indicators (D.5).

- (D.4) In light of this, these maps are perhaps the *oldest* [economic maps<sub>{CSHead}</sub> in the world] **since**<sub>{IN}</sub> they<sub>{PRP}</sub> predate Strabo’s economic maps.

[History of cartography]

- (D.5) The *largest* [project<sub>{CSHead}</sub> at the Canal **since**<sub>{IN}</sub> its<sub>{PRPS}</sub> original construction], the expansion will double its capacity and allow more traffic.

[Panama Canal expansion referendum, 2006]

Once a token has been identified as potential PP-phrase indicator, the ppPhrase

module proceeds to the right, and appends all following tokens that are chunked I-NP to the CSSpan list (again, some exclusions and other minor rules apply, which won't be discussed any further).

Modules II and III, **advPhrase** and **adjPhrase**, consist of a number of simple rules which mainly involve adverbial (I-ADVP and B-ADVP) and adjectival chunks (I-ADJP and B-ADJP), and the Named Entity tags B-DAT and I-DAT. They help to identify cases like (D.6), where the CS span contains the adverbial *entirely*, and (D.7), which includes the adjectival phrase *possible*.

(D.6) During their heyday, tram services covered much of inner London and reached out to the suburbs, assisted by facilities like the Kingsway tramway subway, which enabled the *longest* [tram route **entirely**<sub>{I-ADVP}</sub> within the County of London] to operate: a weekend service between what is now Archway which was then part of Highgate, and Downham via Brockley, 16 miles.

[*Trams in London*]

(D.7) Being raised to the Blood, while exceptionally rare, is the *greatest* [honor **possible**<sub>{I-ADJP}</sub> for one who was born a commoner].

[*Seanchan*]

Postmodifiers of Type IV, non-finite clauses, are recognised via three separate modules: **nonfinClause\_TOVB**, **nonfinClause\_VBG**, and **nonfinClause\_VBN**. These aim to represent *to*-infinitival, gerund-participial, and past-participial non-finite clauses respectively (Huddleston and Pullum, 2002). Module **nonfinClause\_TOVB** succeeds if the focus token has the POS tag TO, and its following token is tagged VB (verb, base form). Once these have been appended to the CSMain list, the module checks the following tokens and adds them to the list if they fulfil certain criteria, e.g. are part of an NP-chunk (D.8).

(D.8) It is the oldest surviving world map from East Asia, and the *oldest* [Asian map<sub>{CSHead}</sub> **to**<sub>{TO}</sub> depict<sub>{VB}</sub> the<sub>{I-NP}</sub> Western<sub>{I-NP}</sub> world<sub>{I-NP}</sub>]

[*Ancient world maps*]

The module created to recognise gerund-participial clauses (**nonfinClause\_VBG**) works in a similar way. Here the focus token has to be tagged VBG (D.9). The lemma

“to be”, however, is excluded from consideration, as it is more likely to express the ISA relation between target and comparison set, as shown in (D.10) .

- (D.9) The Cobble Hill Tunnel and Murray Hill Tunnel in New York City are the world’s *oldest* [railway tunnels<sub>{CSHead}</sub> **lying**<sub>{VBG}</sub> below streets], roofed over in 1850 and the 1850s, respectively.

[*Tunnel*]

- (D.10) The chronological range of the collection spans from the Prehistoric to the Islamic Period with the *largest* [archaeological site collections<sub>{CSHead}</sub>] **being**<sub>{VBG}</sub> Abydos, Amarna, Beni Hasan, Esna and Meroe.

[*World Museum Liverpool*]

Finally, the recogniser for past-participial non-finite clauses (*nonfinClause\_VBN*) deals with cases like (D.11), where the focus token is tagged VBN.

- (D.11) The *most common* [types<sub>{CSHead}</sub> of art<sub>{ppPhrase}</sub> **created**<sub>{VBN}</sub> using models] are figure drawing, figure painting, sculpture and photography.

[*Model (art)*]

Module V aims to detect relative clauses, and uses two different approaches. The first one (**relClause**) attempts to identify what Huddleston and Pullum (2002) call “overt cases”, where the relative clause is introduced by a *wh*-relative (e.g. *who*, *whom*, *whose*, *which*, etc.). The second method (**relClause\_BNP**) aims to find “bare” relative clauses where the anaphoric element is covert (gapped). Due to their structural similarity with overt cases, *that*-relatives are dealt with by the first method.

Identifying relative clauses represents a challenge, as they introduce subordinate clauses which contain at least one noun (the subject of the subordinate clause), which may itself be postmodified, and at least one verb. In addition to the variety of constituents they can contain, there is also the issue of distinguishing non-restrictive from restrictive relative clauses. The pilot study in Chapter 4 showed that while it should theoretically be possible to distinguish between them by the presence or absence of a comma, respectively, in practice this does not always work. As a great deal of world knowledge is needed to disambiguate between the two cases, Module V does not offer a solution for this issue, but simply assumes that commas introduce non-restrictive relative clauses.

For `relClause` to succeed, the focus token needs to be tagged WDT (*wh*-determiner), WP (*wh*-pronoun), or have the lemma “that” (which can also be tagged as IN). After adding the token to the `CSMain` list, the word tokens to the right are investigated, and appended to the list if they conform to a particular pattern, consisting mainly of NP, VP and ADVP chunks. Crucially, as a verb is an obligatory component of a relative clause, the first VP chunk encountered acts as a stopper. This means that all tokens to the left (and including the VP chunk) are appended to the `CSSpan` list, and any I-NP tokens directly following the VP are also added (D.12). If the Identifier encounters a preposition or adjectival chunk during processing, the relative clause is interrupted at that point, and the following tokens are recognised via `ppPhrase` or `adjPhrase`. ADVP chunks, however, are included in the pattern as they can occur before the VP chunk.<sup>2</sup> Example (D.13) shows that the module is also able to identify so-called “nominal” (or independent) relative clauses, which function in some respects like noun phrases. One special restriction is worth noting, which applies to cases of “which” that are preceded by the preposition “of”. They are appended to the `CSMain` list, but then cause the Identifier to stop, as they do not introduce a relative clause (D.14).

(D.12) A similar comparison has been drawn among Go, chess and backgammon, perhaps the three *oldest* [`games`<sub>{CSHead}</sub> **that**<sub>{that}</sub> *still*<sub>{I-ADVP}</sub> *enjoy*<sub>{I-VP}</sub> *worldwide*<sub>{I-NP}</sub> *popularity*<sub>{I-NP}</sub>].

[Go (board game)]

(D.13) The oldest roots of crokinole from the 1860s suggest the British and South Asian games are the *most likely* [`antecedents`<sub>{CSHead}</sub> *of*<sub>{ppPhrase}</sub> *what*<sub>{WP}</sub> *became*<sub>{I-VP}</sub> *crokinole*<sub>{I-NP}</sub>].

[Crokinole]

(D.14) There are several file formats used to store game records, the *most popular*<sub>{CSHead}</sub> [*of which*] is SGF, short for Smart Game Format.

[Go (board game)]

The module for recognising “bare” relative clauses works in a similar way. The challenge here is to recognise the NP that marks the beginning of the relative clause,

<sup>2</sup>The module is not yet able to correctly identify nested relative clauses (such as “*The house that Peter who I know owns*”). However, cases like this are assumed to only occur very rarely in superlative NPs.

which is not straightforward, as the preceding token is also likely to be a noun (D.15). The `relClause` module relies on the chunker to indicate a start of a new NP, and is therefore only triggered if the focused token is chunked B-NP (beginning of NP).<sup>3</sup>

(D.15) The *most common* [forms<sub>{CSHead}</sub> of prosthetics and enhancement<sub>{ppPhrase}</sub> **we**<sub>{B-NP}</sub> see<sub>{I-VP}</sub> in sports today] are prosthetic legs and Tommy John surgery.

[Cyborg]

The last `PostMod` module is a simple detector for NP phrases (or parts of NP phrases) that have not been recognised by any of the previous modules (**npPhrase**).

Crucially, it is added to the processing pipeline *after* the `relClause_BNP` module, as a token chunked B-NP may still be part of the CS span, even if it does not introduce a relative clause.

Finally, the two modules **andMethod** and **commaMethod** aim to tackle coordination, which is generally known to be a difficult problem in NLP. The module `andMethod` is triggered when CS-Identifier encounters the lemma “and”, and its main purpose is to disambiguate between cases like (D.16) and (D.17). In (D.16), “and” describes an NP coordination between “Archbishop” and “Prince-Elector”, and should therefore be included in the CS span (along with the following NP). In (D.17), on the other hand, “and” represents a coordination of two main clauses, and should not be included. The `andMethod` module checks that the token preceding “and” is part of an NP chunk, and includes “and” if its following token is also chunked I-NP (and proceeds to append all following I-NP tokens). However, the module fails for cases where the following token is POS-tagged DT, PRP\$ or PRP, as these are indicators for a separate main clause (as in D.17).

(D.16) Elected in 1307 when he was only 22 years old, Baldwin was the *most important* [Archbishop<sub>{CSHead}</sub> **and** Prince-Elector<sub>{I-NP}</sub> of Trier in the Middle Ages].

[History of Trier]

(D.17) It is one of Bavaria’s oldest and *most popular* [festivals<sub>{CSHead}</sub>] **and** it<sub>{PRP}</sub> has increased to one of the biggest events in Germany.

---

<sup>3</sup>Possessive ‘s, which is usually chunked as B-NP, is treated as I-NP throughout and is therefore excluded from consideration.

[Gäubodenvolksfest]

Like the word “and”, commas can indicate NP coordination, as for example in “the best writer, director, and actor”. What makes such cases more difficult to identify than “and” coordination is the fact that commas also often indicate apposition (D.18).

- (D.18) The entire district contains 14 Israeli settlements, with a total population of 10,000 and two of the *largest* [Palestinian refugee camps in the West Bank], Askar **and** Balata, which comprise about 8% of the total district population.

[Nablus]

To account for comma-based NP coordination, the `commaMethod` module makes the following two assumptions:

1. Appositions usually involve proper nouns.
2. Comma-based NP coordination requires the word “and” as final conjoining element.

Therefore, the method checks if the comma is followed by a series of tokens tagged NN, NNS, or commas (but excluding proper nouns, NNP/NNPS), and only succeeds if it encounters the word “and”.<sup>4</sup>

Sometimes, commas are followed by proper nouns that further modify the CS head and should be included in the CS span. The `commaMethod` contains rules for examples like (D.19), which require the proper noun (here: *Pennsylvania*), to be followed by another comma or full stop.

- (D.19) Reservoir Park is the *largest* [municipal public park<sub>{CSHead}</sub> in Harrisburg<sub>{NNP}</sub>, Pennsylvania<sub>{NNP}</sub>], and occupies approximately 85-acres in the Allison Hill neighborhood of the city.

[Reservoir Park (Harrisburg)]

---

<sup>4</sup>Although no such examples were encountered in the development data, the method has been tested on made up examples.

## D.2 T/CS-Identifier: Results

Table D.1: T/CS-Identifier: Results

N	ID	Sentence	class	T	D	H
1	Altitude sickness	{AMS} is [the] <i>most frequent</i> [ <u>type of altitude sickness</u> encountered].	ISA-1	y	y	n
2	Altitude sickness	{High altitude pulmonary edema (HAPE) and <u>cerebral edema</u> (HACE)} are [the] <i>most ominous</i> [of these symptoms], while acute mountain sickness, retinal haemorrhages, and peripheral edema are the less severe forms of the disease.	ISA-1	n	y	y
3	Altitude sickness	[The] <i>most effective</i> [source of supplemental oxygen at high altitude] are { <u>oxygen concentrators that use vacuum swing adsorption</u> (VSA) technology}.	ISA-1	n	y	y
4	Association football	{Association football}, commonly known as football or soccer, is a team sport played between two teams of eleven players, and is [the] <i>most popular</i> [sport in the world].	ISA-1	y	y	y
5	Association football	[The] <i>most prestigious</i> [international football competition] is {the FIFA World Cup}, held every four years.	ISA-1	y	y	y

Continued on Next Page...

6	Association football	[The world's] <i>oldest</i> [football competition] is {the FA Cup}, which was founded by C. W. Alcock and has been contested by English teams since 1872.	ISA-1	y	y	y
7	Association football	This area has a number of functions, [the] <i>most prominent</i> being {to mark <u>where</u> the goalkeeper may handle the ball and where a penalty foul by a member of the defending team becomes punishable by a penalty kick}.	ISA-1	n	y	y
8	Association football	[The] <i>most complex</i> [of the Laws] is {offside}.	ISA-1	y	y	y
9	Association football	After the World Cup, [the] <i>most important</i> [football competitions] are {the continental championships}, which are organised by each continental confederation and contested between national teams.	ISA-1	y	y	y
10	Association football	[The] <i>most prestigious</i> [competitions in club football] are {the respective continental championships}, which are generally contested between national champions, for example the UEFA Champions League in Europe and the Copa Libertadores de America in South America.	ISA-1	y	y	y
11	Association football	Today the sport is generally known simply as football in countries where {it} is [the] <i>most popular</i> [football code].	ISA-1	y	y	y
12	Dada	A reviewer from the American Art News stated at the time that ”{The Dada philosophy} is [the] <i>sickest</i> , most paralyzing and most destructive [thing that has ever originated from the brain of man].	ISA-1	y	y	y

Continued on Next Page...

13	Dada	A reviewer from the American Art News stated at the time that ”{The Dada philosophy} is [the] sickest, <i>most paralyzing</i> and most destructive [thing that has ever originated from the brain of man].	ISA-1	y	y	y
14	Dada	A reviewer from the American Art News stated at the time that ”{The Dada philosophy} is [the] sickest, most paralyzing and <i>most destructive</i> [thing that has ever originated from the brain of man].	ISA-1	y	y	y
15	Dada	[The] <i>most important</i> [figure in this group] was {Iliazd}, whose radical typographical designs visually echo the publications of the Dadaists.	ISA-1	y	y	y
16	Gonk	Perhaps [the] <i>most famous</i> [gonk in living memory] was {Humpty from Play School}.	ISA-1	y	y	y
17	Gonk	{A gonk in computer-speak} is [the user’s] <i>least favourite</i> [ <u>piece of hardware</u> ].	ISA-1	y	y	n
18	Mana (series)	{The Rabite} has become a sort of mascot for the Mana series, <u>much the same way</u> as the Chocobo represents Final Fantasy, and is [one of its] <i>most recognizable</i> [icons].	ISA-1	n	y	y
19	Mana (series)	{Secret of Mana} is also [the number 6] <i>most remixed</i> [soundtrack on the popular video game music site OverClocked ReMix], with Seiken Densetsu 3 tied at 18.	ISA-1	y	n	y
20	Matryoshka doll	[The] <i>largest</i> [doll] was {that of Fukurokuju} - a happy, bald god with an unusually tall chin - and within it nested the six remaining deities.	ISA-1	y	y	y

Continued on Next Page...

21	Matryoshka doll	Starting with [the] <i>largest</i> , {Mikhail Gorbachev}, then Leonid Brezhnev (Yuri Andropov and Konstantin Chernenko almost never appear due to the short length of their respective terms), then Nikita Khrushchev, Josef Stalin and finally the smallest, Vladimir Lenin.	ISA-1	y	y	y
22	Matryoshka doll	Starting with the largest, Mikhail Gorbachev , then Leonid Brezhnev (Yuri Andropov and Konstantin Chernenko almost never appear due to the short length of their respective terms), then Nikita Khrushchev, <u>Josef Stalin</u> <sub>{T}</sub> and finally [the] <i>smallest</i> , { <u>Vladimir Lenin</u> <sub>{CSMain}</sub> }.	ISA-1	n	y	n
23	New National Theatre	<u>The centre has been praised for its architecture and {state-of the-art modern theatre facilities}</u> , [among the] <i>best</i> [in the world], but criticized for its bureaucratic arts administration.	DEF	n	y	y
24	Oddfellows Playhouse	[The] <i>largest</i> and most active [year-round youth theater in Connecticut], {the Playhouse} is <u>an independent</u> , non-profit organization with an arts/social service mission.	DEF	n	y	y
25	Oddfellows Playhouse	[The] largest and <i>most active</i> [year-round youth theater in Connecticut], {the Playhouse} is <u>an independent</u> , non-profit organization with an arts/social service mission.	DEF	n	y	y
26	Self	[The] first, and <i>most popular</i> , [case] is { <u>the name of the GNU</u> (G NU's N ot U nix) project}.	ISA-1	n	y	y

Continued on Next Page...

27	Vehicular cycling	Advocates of vehicular cycling - such as John Forester, John Franklin and John S. Allen - argue that {cycling in accordance to the vehicular rules of the road} is [the] <i>safest</i> and most effective [means to use a bicycle for transportation].	ISA-1	y	y	y
28	Vehicular cycling	Advocates of vehicular cycling - such as John Forester , John Franklin and John S. Allen - argue that {cycling in accordance to the vehicular rules of the road} is [the] safest and <i>most effective</i> [means to use a bicycle for transportation].	ISA-1	y	y	y
29	Borders of the continents	Although {Greenland} is [the] <i>closest</i> [land to them], they are much closer to Europe than to the North American mainland.	ISA-1	y	y	y
30	Cartography	[The] <i>earliest</i> [known map to date] is {a wall painting of the ancient Turkish city of atal Hyk which has been dated to the late 7th millennium BCE}.	ISA-1	y	y	y
31	Cartography	He also claimed, when considering all aspects of cartography, that ”{map design} is perhaps [the] <i>most complex</i> .	ISA-1	y	y	y
32	Kyzyl Kum	{The Kyzyl Kum} (Uzbek: Qizilqum), also called Qyzylqum, is [the] <i>11th largest</i> [desert in the world].	ISA-1	y	y	y
33	Meander	The stochastic theory can take many forms but [one of the] <i>most general</i> [statements] is { <u>that of Scheidegger</u> : “The meander train is assumed to be the result of the stochastic fluctuations of the direction of flow due to the random presence of direction-changing obstacles in the river path.”}	ISA-1	n	y	y

Continued on Next Page...

34	Nicopolis	Besides the Acropolis, [the] <i>most conspicuous</i> [objects] are { <u>two theatres</u> (the larger with 77 rows of seats) and an aqueduct which brought water to the town from a distance of 27 miles}.	ISA-1	n	y	y
35	Population density	While { <u>the arithmetic density</u> } is [the] <i>most common</i> [way of measuring population density], several other methods have been developed which aim to provide a more accurate measure of population density over a specific area.	ISA-1	n	y	y
36	St Kilda, Scotland	[The] <i>largest</i> [island] is {Hirta}, whose sea cliffs are the highest in the United Kingdom.	ISA-1	y	y	y
37	St Kilda, Scotland	The largest island is Hirta, <u>whose</u> { <u>sea cliffs</u> } are [the] <i>highest</i> [in the United Kingdom].	ISA-1	n	y	y
38	St Kilda, Scotland	At 670 ha (1,656 acres) in extent, {Hirta} is [the] <i>largest</i> [ <u>island in the group</u> ] <u>and</u> comprises more than 78% of the land area of the archipelago.	ISA-1	y	y	n
39	St Kilda, Scotland	[The] <i>highest</i> [point in the archipelago], {Conachair} ('the beacon') at 430 metres (1,411 ft), is <u>on</u> Hirta, immediately north of the village.	DEF	n	y	y
40	St Kilda, Scotland	[The] <i>oldest</i> [building] is { <u>an underground passage with two small annexes</u> called Taigh an t-Sithiche (house of the faeries) which dates to between 500 BC and 300 AD}.	ISA-1	n	y	y
41	St Kilda, Scotland	{Christ Church}, in the site of the graveyard at the centre of the Village, was in use in 1697 and was [the] <i>largest</i> , but this thatched-roof structure was too small to hold the entire population, and most of the congregation had to gather in the churchyard during services.	ISA-1	y	y	y

Continued on Next Page...

42	Venezuela	[Venezuela] is [among the] <i>most urbanized</i> [countries in Latin America]; the vast majority of Venezuelans live in the cities of the north, especially in the capital Caracas which is also the largest city.	ISA-1	y	y	y
43	Venezuela	Venezuela is among the most urbanized countries in Latin America; the vast majority of Venezuelans live in the cities of the north, especially in {the capital <i>Caracas</i> } which is also [the] <i>largest</i> [city].	ISA-1	n	y	y
44	Venezuela	The discovery of massive oil deposits, totaling some 400 million barrels, during World War I prompted <u>an economic boom that lasted into the 1980s</u> ; by 1935, {Venezuela's per capita GDP} was [Latin America's] <i>highest</i> , and globalization and heavy immigration from Southern Europe and poorer Latin American countries markedly diversified Venezuelan society.	ISA-1	n	y	y
45	Venezuela	[The] <i>highest</i> [judicial body] is {the Supreme Tribunal of Justice} or Tribunal Supremo de Justicia, whose magistrates are elected by parliament for a single twelve-year term.	ISA-1	y	y	y
46	Venezuela	{Pico Bolvar}, [the nation's] <i>highest</i> [point] at 4,979 metres (16,335 ft), lies in this region.	ISA-1	y	y	y
47	Venezuela	To the south, the dissected Guiana Highlands is home to the northern fringes of <u>the Amazon Basin</u> and { <u>Angel Falls</u> }, [the world's] <i>highest</i> [waterfall].	ISA-1	n	y	y
48	Venezuela	Notable mammals include the giant anteater, jaguar, and {the capybara}, [the world's] <i>largest</i> [rodent].	ISA-1	y	y	y

Continued on Next Page...

49	Venezuela	{Carlos Ral Villanueva} was [the] <i>most important</i> [Venezuelan architect of the modern era]; he designed the Central University of Venezuela, (a World Heritage Site) and its Aula Magna.	ISA-1	y	y	y
50	Venezuela	{Baseball} is [Venezuela's] <i>most popular</i> [sport], although football (soccer), spearheaded by the Venezuela national football team, is gaining influence.	ISA-1	y	y	y
51	Frontal lobe epilepsy	{Frontal lobe epilepsy} is [the] <i>second most common</i> [type of epilepsy], after temporal lobe epilepsy.	ISA-1	y	y	y
52	Harm reduction	Others think [the] <i>best</i> [option] would be {some degree of decriminalization}, such as allowing the possession of small amounts of cannabis and possibly its cultivation for personal use, while concentrating law-enforcement resources on more serious crimes, e.g. crimes that have victims instead of an individual breaking a law of prohibition.	ISA-1	y	y	y
53	Hip replacement	Surgery costs vary from country to country, with { <u>the US</u> } typically being [among the] <i>highest</i> [markets], and countries like Thailand, Cuba and Argentina, among the lowest.	DEF	n	y	y
54	Hip replacement	Surgery costs vary from country to country, with <u>the US</u> typically being among the highest markets, and {countries like Thailand, Cuba and Argentina}, [among the] <i>lowest</i> .	DEF	n	y	y

Continued on Next Page...

55	History of psychotherapy	[The] <i>earliest</i> [recorded approaches] were {a <u>combination of religious</u> , magical and/or medical perspectives}.	ISA-1	n	y	y
56	History of psychotherapy	[The] <i>earliest</i> [recorded approaches] were {a <u>combination of religious</u> , magical and/or medical perspectives}.	ISA-1	n	y	y
57	History of psychotherapy	Perhaps [the two] <i>biggest</i> have been { <u>Systems Therapy</u> which focuses on family and group dynamics and Transpersonal psychology}, which focuses on the spiritual facet of human experience.	ISA-1	n	y	y
58	Infectious diarrhea	[The] <i>most common</i> [ <u>organisms found</u> ] are { <u>Campylobacter</u> (from animal products), Salmonella (also often from animal foodstuffs), Cryptosporidium (ditto), and Giardia lamblia (lives in water)}.	ISA-1	n	y	n
59	Infectious diarrhea	[The] <i>most common</i> is {pinworm} (mostly a nuisance rather than a severe medical illness).	ISA-1	y	y	y
60	Infectious diarrhea	{Salmonella} is [the] <i>most common</i> [persistent bacterial organism in humans].	ISA-1	y	y	y
61	Long Slow Distance	Kurt Wilkens is sceptical that except for marathon and triathlon training, {LSD} is not [the] <i>best</i> [way for maximizing athletic performance].	ISA-1	y	y	y

Continued on Next Page...

62	Massage	{ <u>Heinrich Himmler</u> }, <u>commander of the Schutzstaffel</u> (SS) and [one of the] <i>most powerful</i> [men in Nazi Germany] might have lost faith in German victory due to his discussions with his masseurs Felix Kersten and Walter Schellenberg.	ISA-1	n	y	y
63	Underweight	[The] <i>most common</i> [ <u>cause of a person</u> being underweight] is primarily {malnutrition caused by the unavailability of adequate food}, which can run as high as 50% in parts of sub-Saharan Africa and south Asia.	ISA-1	y	y	n
64	Underweight	[The] <i>most immediate</i> [problem with underweight] is { <u>that it might be secondary to</u> , and/or symptomatic of, an underlying disease}.	ISA-1	n	y	y
65	British Peer (ship)	Originally {she} was [one of the] <i>fastest</i> [vessels in her class] but alterations to increase her tonnage by lengthening her hull by 32 feet in 1878 completely spoiled her sailing powers and she was never as fast again.	ISA-1	y	y	y
66	History of biology	{The philosopher Aristotle} was [the] <i>most influential</i> [scholar of the living world from classical antiquity].	ISA-1	y	y	y
67	History of biology	{Pliny the Elder} was also known for his knowledge of plants and nature, and was [the] <i>most prolific</i> [compiler of zoological descriptions].	ISA-1	y	y	y
68	History of biology	[The] <i>most significant</i> [evolutionary theory before Darwin's] was {that of Jean-Baptiste Lamarck}; based on the inheritance of acquired characteristics (an inheritance mechanism that was widely accepted until the 20th century), it described a chain of development stretching from the lowliest microbe to humans.	ISA-1	y	y	y

Continued on Next Page...

69	History of biology	In the 1970s Stephen Jay Gould and Niles Eldredge proposed the theory of punctuated equilibrium which holds that {stasis} is [the] <i>most prominent</i> [feature of the fossil record], and that most evolutionary changes occur rapidly over relatively short periods of time.	ISA-1	y	y	y
70	History of biology	{The Human Genome Project} - [the] <i>largest</i> , most costly [ <u>single biological study</u> ever undertaken] - began in 1988 under the leadership of James D. Watson, after preliminary work with genetically simpler model organisms such as E. coli, S. cerevisiae and C. elegans.	DEF	-	y	n
71	History of biology	{The Human Genome Project} - [the] <i>largest</i> , <i>most costly</i> [ <u>single biological study</u> ever undertaken] - began in 1988 under the leadership of James D. Watson, after preliminary work with genetically simpler model organisms such as E. coli, S. cerevisiae and C. elegans.	DEF	-	y	n
72	History of Mauritius	However, <u>it</u> was only from 1735, with the arrival of [the] <i>most illustrious</i> [of French governor], {Mah de La Bourdonnais}, that the “Isle de France” started developing effectively.	DEF	n	y	y
73	History of Mauritius	[One of the] <i>most important</i> [events] was {the abolition of slavery in 1835}.	ISA-1	y	y	y
74	Muhammad of Ghor	Upon his death, {Qutb-ud-din Aybak}, [Muhammad Ghori’s] <i>most capable</i> [general], who had started of by sacking Ayodhya in 1193 A.D., took control of Muhammad’s Indian conquests and declared himself the first Sultan of Delhi thus establishing Sultanate of Delhi in 1206 CE.	DEF	-	y	y

Continued on Next Page...

75	Oskars Kalpaks	After his death, <u>Kalpaks</u> was posthumously awarded [Latvia's] <i>highest</i> [military award], {'Order of Lacplesis I III Class'}	DEF	n	y	y
76	Priestly Blessing	{This} is [the] <i>oldest</i> [known Biblical text that has been found]; amulets with these verses written on them have been found in graves in dating from the First Temple Period, and are now in the Israel Museum, Jerusalem.	ISA-1	y	y	y
77	Supreme Federal Court (Brazil)	{The Supreme Federal Court} (in Portuguese Supremo Tribunal Federal, or simply STF), is the Brazilian Supreme Court, [the] <i>highest</i> [court of law of the Federative Republic of Brazil].	ISA-1	y	y	y
78	War of the Succession of Flanders and Hainault	When Baldwin IX, Count of Flanders and Hainault, left on the Fourth Crusade in 1202, he left his western domains under [his] <i>eldest</i> [ <u>daughter</u> ] { <u>Joanna</u> }.	DEF	-	y	n
79	Animal	[Their] <i>closest</i> [known living relatives] are {the choanoflagellates}, collared flagellates that have a morphology similar to the choanocytes of certain sponges.	ISA-1	y	y	y
80	Animal	[The] <i>largest</i> [animal phylum] belongs here, {the Arthropoda}, including insects, spiders, crabs, and their kin.	DEF	-	y	y
81	Animal	The ecdysozoans also include {the Nematoda} or roundworms, [the] <i>second largest</i> [animal phylum].	ISA-1	y	y	y

Continued on Next Page...

82	Animal	Flatworms are acoelomates, or lack a body cavity, as do [their] <i>closest</i> [relatives], {the microscopic Gastrotricha}.	ISA-1	y	y	y
83	Animal	[The] <i>most prominent</i> are {the Rotifera} or rotifers, which are common in aqueous environments.	ISA-1	y	y	y
84	Animal	The Lophotrochozoa include [two of the] <i>most successful</i> [animal phyla], {the Mollusca and Annelida}.	DEF	-	y	y
85	Medicine	[The] <i>earliest</i> [type of medicine in most cultures] was { <u>the use of empirical natural resources like plants</u> (herbalism), animal parts and minerals}.	ISA-1	n	y	y
86	Psychology	[One of the] first, and <i>smallest</i> , [models] was {that of Hans Eysenck}, which had three dimensions: extroversion - introversion, neuroticism - emotional stability, and psychoticism.	ISA-1	y	y	y
87	Psychology	{Industrial and organizational psychology} (I/O) is [among the] <i>newest</i> [fields in psychology].	ISA-1	y	y	y
88	Ranunculaceae	Numerically [the] <i>most important</i> [genera] are { <u>Ranunculus</u> (600 species), Delphinium (365 species), Thalictum (330 species), Clematis (325 species), and Aconitum (300 species)}.	ISA-1	n	y	y
89	Subarctic climate	Even though the diversity may be low, numbers are high, and the {taiga (boreal) <u>forest</u> } is [the] <i>largest</i> [forest biome on the planet], with most of the forests located in Russia and Canada.	ISA-1	n	y	y

Table D.2: Target errors

N	ID	Sentence	Class	P.Class	THead
13	wsj04 485	The candidates have been crisscrossing this huge country of 145 million people , holding rallies and televised debates in hope of being elected to <u>what must</u> <sub>{THead}</sub> be [one of the world 's] <i>most thankless</i> [political jobs] : {trying to drag Brazil out of its economic and social mess} .	ISA-1	DEF	back-off
26	wsj04 813	But {Mr. Iacocca 's <u>remarks</u> <sub>{THead}</sub> } are [the] <i>most specific</i> [indication to date of how many plants could be in jeopardy] .	ISA-1	ISA-1	xcompos
31	wsj04 994	First Boston 's merchant banking risks mounted last month as <u>highly leveraged</u> { <u>Campeau Corp.</u> <sub>{THead}</sub> }, [First Boston 's] <i>most lucrative</i> [client of the decade] , was hit by a cash squeeze and the high-risk junk bond market tumbled .	ISA-1	ISA-1	subjapp
33	wsj04 1011	The spokeswoman said the move is n't directly a response to Quotron 's loss of [its two] <i>biggest</i> [customers] , {Merrill Lynch & <u>Co.</u> <sub>{THead}</sub> and <u>American Express Co.</u> 's <u>Shearson Lehman Hutton Inc.</u> } , to Automated Data Processing Inc. earlier this year .	ISA-1	ISA-1	objapp

Continued on Next Page...

50	wsj04 1605	{ <u>The \$</u> <sub>{THead}</sub> 500 million in notes} , [the] <i>largest</i> [chunk of Western Union 's \$ 640 million in long-term debt] , stems from the company 's major restructuring in December 1987 .	ISA-1	ISA-1	subjapp
52	wsj04 1619	Meanwhile , the index of [the 100] <i>biggest</i> [non-financial stocks] , { <u>the Nasdaq</u> <sub>{THead}</sub> 100} , gained 0.47 to 438.15 .	ISA-1	ISA-1	objapp
55	wsj04 1713	Dr. Vogelstein next turned his attention to {colon cancer} , [the] <i>second biggest</i> [cancer killer in the U.S.] after lung cancer .	ISA-1	FREE	(No CSHead)
61	wsj04 1960	At least two rival applications are expected to emerge in coming months , including <u>one</u> <sub>{THead}</sub> <u>from</u> { <u>TransCanada PipeLines Ltd.</u> } , [Canada 's] <i>largest</i> [natural gas pipeline operator] .	ISA-1	ISA-1	objapp
78	wsj10 7	[The] <i>most volatile</i> [form of program trading] is { <u>index arbitrage</u> <sub>{THead}</sub> } = <u>the rapid-fire</u> , computer-guided buying and selling of stocks offset with opposite trades in stock-index futures and options .	ISA-1	ISA-1	subjpos
80	wsj10	At a time when Jon Levy should be planning the biggest spring season in his dress company 's 17 years , <u>his work day</u> <sub>{THead}</sub> is studded with intense moments of concern about [one of his] <i>biggest</i> [customers] , {Campeau Corp.}	ISA-1	DEF	back-off
85	wsj10 182	National Intergroup , which owns 50 % of [the nation 's] <i>sixth largest</i> [steel-maker] – {National Steel Corp.} – posted net income for the fiscal second-quarter of \$ 8.6 million , or 33 cents a share , compared with a net loss of \$ 50.3 million .	ISA-1	FREE	(No CSHead)

Table D.3: CS errors: DLA wrong, CS-Identifier correct

N	ID	Sentence	Class
1	wsj04 29	The <i>latest</i> [ <b>10-year notes</b> ] ended at about 100 16 / 32 to yield 7.90 % , compared with 100 11/32 to yield 7.93 % on Friday .	IDIOM
6	wsj04 231	Like Mr. Geffen 's arrangement , the venture gives Mr. Azoff a link to [ <b>the world 's</b> ] <i>largest</i> and most successful [ <b>record distributor</b> ] ; in the U.S. alone , Warner has a 40 % share of the market , about double its nearest competitor , Sony Corp. 's CBS Records .	DEF
7	wsj04 231	Like Mr. Geffen 's arrangement , the venture gives Mr. Azoff a link to [ <b>the world 's</b> ] <i>largest</i> and <i>most successful</i> [ <b>record distributor</b> ] ; in the U.S. alone , Warner has a 40 % share of the market , about double its nearest competitor , Sony Corp. 's CBS Records .	DEF
10	wsj04 257	For instance , sales of treadmills , exercise bikes , stair climbers and the like are expected to rise 8 % to about \$ 1.52 billion this year , according to the National Sporting Goods Association , which sees the home market as one of the <i>hottest</i> [ <b>growth areas for the 1990s</b> ] .	ISA-2
19	wsj04 526	WHICH IS the <i>best</i> <u>medicine for runaway health costs</u> : <u>competition</u> or <u>regulation</u> ?	DEF
30	wsj04 973	The seasonally adjusted Dodge Index reached 175 in September , [ <b>its</b> ] <i>highest</i> [ <b>level this year</b> ] , from 162 in August .	ISA-1

Continued on Next Page...

33	wsj04 1011	The spokeswoman said the move is n't directly a response to Quotron 's loss of <b>[its] two biggest [customers]</b> , Merrill Lynch & Co. and American Express Co. 's Shearson Lehman Hutton Inc. , to Automated Data Processing Inc. earlier this year .	ISA-1
35	wsj04 1154	Toshiba Corp. busted open that sector this summer with a notebook-sized machine that retails for less than 200,000 yen ( under \$ 1,500 ) – one of the smallest , <i>cheapest</i> <b>[PCs available in the country]</b> .	DEF
36	wsj04 1154	Toshiba Corp. busted open that sector this summer with a notebook-sized machine that retails for less than 200,000 yen ( under \$ 1,500 ) – one of the <i>smallest</i> , cheapest <b>[PCs available in the country]</b> .	DEF
38	wsj04 1180	From a reading of the somewhat scant English-language medical literature on RU-486 , the French abortion pill emerges as one of the <i>creepiest</i> <b>[concoctions around]</b> .	ISA-2
41	wsj04 1272	The decision by Merrill , the nation 's largest securities firm , represents the <i>biggest</i> <b>[retreat yet from program trading]</b> .	DEF
42	wsj04 1273	Merrill has been the <i>fourth-biggest</i> <b>[stock-index arbitrage trader on the Big Board this year]</b> , executing an average of 18.1 million shares a month in such trades , or about one million shares a day .	ISA-1
44	wsj04 1284	Just a week ago , Mr. Carpenter staunchly defended index arbitrage at Kidder , the <i>most active</i> <b>[index-arbitrage trading firm on the stock exchange this year]</b> .	ISA-1
58	wsj04 1798	But the transaction is just <b>[Mr. Peladeau 's] latest [step in a larger design]</b> : <u>to build Quebecor through acquisitions into an integrated paper , publishing and printing concern with a reach throughout North America .</u>	IDIOM

Continued on Next Page...

74	wsj04 1346	About 140 Mayan , Aztec , Mixtec and Zapotec objects , including some of [ <b>Mexico 's</b> ] <i>best-known</i> [ <b>archaeological treasures</b> ] , were taken .	DEF
76	wsj04 1750	In any case , says Dr. Minna of the National Cancer Institute , We 're witnessing the discovery of one of the <i>most important</i> [ <b>steps in the genesis of cancer</b> ] .	DEF
77	wsj10 2	The move is the <i>biggest</i> [ <b>salvo yet in the renewed outcry against program trading</b> ] , with Kemper putting its money – the millions of dollars in commissions it generates each year – where its mouth is .	ISA-1
88	wsj10 229	Provigo was the <i>most active</i> [ <b>industrial stock on the Montreal Exchange</b> ] , <u>where it closed at C\$ 9.75 ( US\$ 8.32 ) , up 75 Canadian cents .</u>	ISA-1

# Appendix E

## Appendix to Chapter 7

### E.1 Discussion of SUP-Classifier results on Hu and Liu (2004)'s corpus of customer reviews

This section discusses the errors produced by SUP-Classifier on the customer review data. The confusion matrix in Table E.1 shows the Classifier's prediction for each of the classes.

Row	Class	IDIOM	PP	PROP	ADV	FREE	INDEF	ISA-1	ISA-2	DEF
1	<b>IDIOM</b>	<b>4</b>	0	0	0	0	3	0	0	4
2	<b>PP</b>	0	<b>26</b>	1	0	0	0	0	0	0
3	<b>PROP</b>	0	0	<b>44</b>	0	0	2	0	0	0
4	<b>ADV</b>	0	0	1	<b>6</b>	0	0	1	0	2
5	<b>FREE</b>	0	0	0	2	<b>1</b>	0	0	0	0
6	<b>INDEF</b>	0	0	1	0	2	<b>12</b>	0	0	0
7	<b>ISA-1</b>	0	0	1	2	0	1	<b>53</b>	0	5
8	<b>ISA-2</b>	0	0	0	0	0	0	0	<b>1</b>	7
9	<b>DEF</b>	0	0	0	0	0	0	2	2	<b>39</b>

Table E.1: Confusion matrix

IDIOM-Identifier succeeded in indentifying four superlatives with the help of the idiom lexicon `IdiomLex` with 100% precision. Its recall, however, is considerably lower at only 36.7%. The confusion matrix shows that three IDIOM instances were falsely classified as INDEF, and another four as DEF. While IDIOM-Identifier

failed to recognise cases like and because they are missing from `IdiomLex` (and Chapter 5.3.5 already notes that the phrase “for the most part” should be included), example illustrates a problem that is specific to Hu&Liu’s data set: It seems that their corpus has not only been tokenised but all capital letters have been replaced by small ones. This means that proper names spelt usually spelt with capital letters (e.g. “Best Buy”, which clearly refers to a shop) are not recognised by `IDIOM-Identifier`’s `CapsTest` method (described in Chapter 5.3.3.1). This, however, may also be a general problem within this genre, as reviewers often disregard capitalisation and consistently use small letters (apart from possibly the first word in a sentence).

- (E.1) *best* regards sadoun satellite sales
- (E.2) for the *most* part this is a good dvd player .
- (E.3) short warranty ( 3 months , so i got the extended warranty at *best* buy ) ,  
occasional freeze up ( easily fixed by popping out battery ) , somewhat  
flimsy navigation wheel .

The only error produced by `PP-Identifier` is due to incorrect tokenisation (E.6). Otherwise, this instance would have been recognised by the first PP rule described in Chapter 5.3.3.2.

- (E.4) i end up unplugging it from the wall for a few hours and when i try again it  
will work fine.at least until this problem arises again .

`PROP-Identifier` also has very good recall, with only two instances unrecognised. The first one, which is subsequently classified as `INDEF`, is caused by failure of Method III (cf. Chapter 5.3.3.3). While the tag of the word following `B-SUP` is `JJ`, its length is not  $\geq 6$ , which causes the Classifier to consider it as an analytical superlative. The second instance is not recognised because *most* has not received the supertag `N`, which usually identifies such cases as proportional quantifiers.

- (E.5) battery life , the drawback to *most* digital cameras , to quote k. reeves .
- (E.6) the t610 can also compose music , a feature which *most* wo n’t use .

As Table 7.4 shows, `ADV-Identifier` does not work as well on the customer review data compared to the results for `TextWiki`, with four out of 10 instances not recognised. One of `ADV-Identifier`’s problems with more colloquial data is that hyphenation is treated very inconsistently. For example, (E.7) was missed by `ADV-Identifier`’s “compound search” method (Chapter 5.3.3.4) because there is no hyphen between *finest* and *looking*. Example (E.8), on the other hand, was falsely labelled as `ADV` because the Identifier recognised it as compound due to the hyphen.

(E.7) without a doubt the *finest* looking apex dvd player that i 've seen .

(E.8) the little digital elphs were the best-designed , *most*-elegant cameras until the g3 came along and squashed them .

Of three FREE instances altogether, two were misclassified as ADV in the previous step of the clasification pipeline. The errors, however, are repetitions of each other, with (E.9) being a review title and (E.10) being the first sentence in the review. They could therefore count as one single mistake, caused by ADV-Identifier misinterpreting the word “to” as infinitival marker indicating a relation involving a verb.

(E.9) this camera is *closest* to perfect than any otherdigicam

(E.10) this camera is *closest* to perfect than any other digicam that is 4-5 megapixel .

The errors in the INDEF class are largely due to problems involving abbreviated language. For example, in (E.11), SUP-Classifier recognises the superlative as proportional quantifier because “most” is not preceded by “the”, but occurs as first word in an (incomplete) sentence and is followed by a noun (NN). Even more problematic are cases like (E.12), which occur within a review title. Due to its lack of context and “headline” status classification of such instances as INDEF is debatable.

(E.11) *most* bang for the buck. i recieved my 2600 4 days ago and feel that this may be the best \$ 50 i have ever spent .

(E.12) *best* of the best

Recall of the last Identifier in the pipeline, DEF, is very good, with only four out of 43 instances not identified. These were classified as ISA-1 (E.13 and E.14) and ISA-2 (E.15 and E.16).

(E.13) the *highest* optical zoom pictures are perfect .

(E.14) for a price that 's still less than even the *lowest* level ipod i was able to get this 40gb monster , and the best part is it works as great as it was advertised to and then some .

(E.15) overall , the g3 delivers what must be considered the *best* image quality of any current > 4 megapixel digicams , from a detail , tonal balance and color response point of view .

(E.16) i spent a lot of time comparing different cameras , and i realized that there is not such thing as the *best* digital camera .

Table E.2: T-Identifier Results (Hu&amp;Liu data)

File	Sentence	Features	T	D	H
Apex	{this} is by far [the] <i>nicest</i> [one] , in so many ways .	dvd player[+3][p]	y	y	-
Apex	most bang for the buck. i recieved my 2600 4 days ago and feel that {this} may be [the] <i>best</i> [\$ 50 i have ever spent] .	[t]	n	y	n
Apex	the progressive scan option can be turned off easily by {a button on the remote control} which is [one of the] <i>simplest</i> and <i>easiest</i> [remote controls i have ever seen or used] .	progressive scan[+2], remote[+3]	n	y	<b>n</b>
Apex	the progressive scan option can be turned off easily by {a button on the remote control} which is [one of the] <i>simplest</i> and <i>easiest</i> [remote controls i have ever seen or used] .	progressive scan[+2], remote[+3]	by a button on	y	<b>n</b>
Apex	{this} is [the] <i>best</i> [dvd player i 've purchased] .	player[+3]	y	y	n
Apex	[the] <i>most recent</i> is {that it will stop responding to the on/off button} .	on/off button[-2]	y	y	y
Apex	i think , {apex} is [the] <i>best</i> [dvd player you can get for the price] .	dvd player[+3]	y	y	y

Continued on Next Page...

Apex	what can i say , i bought the ad600 2 years ago and {it} was probly [the] <i>best</i> [ <u>dvd player ever</u> ] , <b>quality design</b> , <u>stable</u> , plus it uses a standard ide interface so you can use an old dvd drive from your computer if the player dies .	dvd player[+3]	n	y	n
Apex	but , dollar for dollar , {this dvd player} is probably [the] <i>best</i> [out there] .	player[+3]	y	y	y
Apex	{apex} is [the] <i>best</i> {cheap quality brand for dvd players} .	apex[+3]	y	y	y
Apex	i bought the apex ad-2600 because of the ability to play jpegs and it seemed like {it} was [the] <i>most</i> [ <u>bang for the buck</u> ] .	-	y	y	n
Canon	i 'm easily enlarging pictures to 8 1/2 x 11 with no visable loss in picture quality and not even using [the] <i>best</i> [ <u>possible setting</u> as yet] ( {super fine} ) .	picture quality[+1]	-	y	n
Canon	<u>overall</u> {it} is [the] <i>best</i> [camera on the market] .	camera[+3]	n	y	y
Canon	fortunately , the g3 has a hot shoe that can take an external flash , and fortunately for me , i already owned a [one of the] <i>highest</i> [quality external flashes that money can buy] : {the metz 60 ct-4} .	external flash hot shoe[+1]	-	y	y
Canon	i did a good month 's worth of research before buying this over other similar priced digital cameras , and {this} is [the] <i>best</i> [buy for the buck] .	-	y	y	y
Canon	{this} is quite simply [the] <i>best</i> [you can ask for] .	camera[+3][p]	y	y	y
Canon	{the Canon} is perhaps [the] <i>best</i> [4mp camera out there] .	4mp camera[+2]	y	y	y
Canon	{the little digital elphs} were [the] best-designed , <i>most-elegant</i> [ <u>cameras until the g3</u> came along and squashed them] .	-	y	y	n
Canon	{this} is by far [the] <i>finest</i> [ <u>camera in its price and category</u> i have ever used] .	camera[+2]	y	y	n

Continued on Next Page...

Canon	all-in-all , i believe {this} is arguably [the] <i>best</i> [non-slr digital camera on the market] .	camera[+3]	y	y	y
Canon	simply , {the Canon} is [the] <i>best</i> [digital camera out there today at this price point] .	digital camera[+3]	y	y	y
Canon	even with these shortcomings , i still think {it} is [the] <i>best</i> [digital camera available under \$ 1200] .	camera[+3]	y	y	y
Creative	{it} is [the] <i>most</i> [bang-for-the-buck <u>out there</u> ] .	bang-for-the-buck[+3][u]	y	y	n
Creative	i chose this one because from what i read , {it} was [the] <i>best</i> [deal for the money] .	deal[+3]	y	y	y
Creative	i 've had <u>it for about 2 weeks and</u> {it} 's [the] <i>best</i> [ <u>player i</u> 've seen , and used , ever] .	player[+3]	n	y	n
Creative	it has several play options , [one of the] <i>best</i> is {an option to shuffle} , based on all or a certain playlist .	play option[+1]	y	y	y
Creative	{this} was [the] <i>closest</i> [alternative] and best buy for this type .	-	y	y	y
Creative	{this} was [the] closest alternative and <i>best</i> [buy for this type] .	-	y	y	y
Creative	in fact , my boyfriend is now going to invest into one of these suckers as well ... now he realizes that {this} is [the] <i>best</i> [ <u>choice over his current archos</u> and the ipod he was eyeing] .	-	y	y	n
Creative	compared to everything else in this category , {this} is most definately [the] <i>best</i> [bang for the buck] .	player[+3]	y	y	y

Continued on Next Page...

Creative	<u>software</u> - {music match jukebox} is n't [the] <i>greatest</i> , the search funtion is n't fast even when accessing it with the hotkey shortcut .	software[-1]	n	y	y
Creative	2 . {the scroll button} is n't [the] <i>best</i> , as it sometimes can be hard to select .	scroll[-2]	y	y	y
Creative	i think { <b>this</b> } is [it 's] <i>biggest</i> [flaw] .	navigation[-3][p]	n	<b>n</b>	y
Creative	{the headphones} are n't [the] <i>best</i> , but you can really expect much out of the small headphones you insert in the ear cavity .	headphone[-1]	y	y	y
Creative	before asking for the zen xtra specifically for christmas i researched many different mp3 players but finally decided on this one , and now that i have it i could n't be happier and i 'm <u>convinced</u> {it} 's [the] <i>best</i> [one out there] .	player[+3]	n	y	-
Creative	perhaps [the] <i>weirdest</i> was { <u>one of bruce springsteen 's cds from the tracks set</u> being classified as goth rock} .	-	n	y	y
Creative	considering the price and specs , {this} might be [the] <i>best</i> [buy out there for a large memory mp3 player] .	player[+2]	y	y	y
Creative	{the construction of the player} is [the] <i>cheesiest</i> [i have ever seen] – the front panel refused to clip in correctly , leaving a noticeable gap between the panel and base of the player .	construction[-3]	y	y	n
Creative	i loved this product when i first got it - great sound , easy to use ( even though i agree that {the software} was n't [the] <i>best</i> ) and held enough songs to suit any mood .	product[+3], sound[+3], use[+2]	y	y	y
Creative	[the] <i>biggest</i> [drawback that people have about the zen xtra] is {the software} .	-	y	y	y

Continued on Next Page...

Creative	[my] <i>biggest</i> [complaint] is { <u>the battery life</u> or lack there of} .	battery life[-3]	n	y	y
Creative	[the] one and <i>most major</i> <u>thing that i was dissapointed with</u> , in this player] , is {the battery life} .	battery life[-3]	y	y	n
Creative	possibly [the] <i>biggest</i> [advantage this player has] ( if price was n't enough ) is {the fact that the user can replace the battery him / herself by opening the lid and popping in the new battery} .	battery[+3]	y	y	y
Creative	for a price that 's still less than even the lowest level ipod i was able to get this 40gb monster , and [the] <i>best</i> [part] is { <u>it works as great as it was advertised to and then some</u> } .	price[+2]	n	y	y
Creative	honestly , [the] <i>best</i> [players without any flaws] were {the goddam cd players} .	-	y	y	y
Nikon	{nikon 4300} , <u>i feel</u> , is [the] <i>best</i> [camera out there for the features and price] .	camera[+3]	n	y	y
Nikon	if you do any research into digital cameras , you 'll quickly find that {this camera} is just about [the] <i>best</i> [value out there] .	-	y	y	y
Nikon	<u>in my opinion</u> {it} 's [the] <i>best</i> [ <u>camera for the money if you</u> 're looking for something that 's easy to use] , small good for travel , and provides excellent , sharp images .	camera[+3]	n	y	n
Nokia	[the two] <i>biggest</i> [things] is { <u>the excellent working speakerphone</u> “ <u>unlike the nokia 3650</u> ” and the superb reception nokia is known for in the gsm phones they make} ...	speakerphone[+3], reception[+2]	n	y	y

Continued on Next Page...

Nokia	while going for a cellphone , i was fully determined to buy a nokia only as {it} is [the] <i>best</i> [in cellphones] .	nokia[+3]	y	y	y
Nokia	and really {this} is [the] <i>best</i> [phone one could have] .	phone[+3]	y	y	y
Nokia	[the] <i>most important</i> [thing for me] is {sound quality} .	-	y	y	y
Nokia	of all these phones {the motorolla and <u>the sony ericsson phones</u> } are [the] <i>smallest</i> .	-	n	y	y
Nokia	{the 6610} is the actually [the] <i>longest</i> but they are all very small .	-	y	y	y
Nokia	{this} is [the] <i>best</i> [ <u>phone i</u> have seen] .	phone[+3]	y	y	n
Nokia	i just purchased this phone and i think {this} is [the] <i>coolest</i> [ <u>phone i ever</u> had] .	phone[+3]	y	y	n
Nokia	in my opinion [the] <i>worst</i> [issue on this phone] is {the side-mounted volume control} .	volume control[-3]	y	y	y
Nokia	<u>overall</u> {this} is [the] <i>best</i> [ <u>phone i</u> have ever owned] .	phone[+2]	n	y	n
Nokia	now { <u>my little 6610</u> } is [the] <i>most coveted</i> [ <u>item at school</u> ] !	-	n	y	y
Nokia	{this} is probably [your] <i>best</i> [ <u>bet if you</u> are looking for a phone in this price range , or like me , do not have the patience to deal with annoying flip phones] .	phone[+3]	y	y	n
Nokia	<u>best phone out there</u> ... a+ not only is {this} [the] <i>best</i> [phone out there] but is cheap ( free ! ! ) .	[t]	n	y	y
Nokia	{this} is [one of the] <i>niciest</i> [ <u>phones nokia</u> has made] .	phone[+3]	y	y	n

# Bibliography

- Ahn, K., Alex, B., Bos, J., Dalmas, T., Leidner, J. L., and Smillie, M. (2004). Cross-lingual Question Answering using off-the-shelf Machine Translation. In *Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004*.
- Bauer, L. (1983). *English Word-Formation*. Cambridge University Press, Cambridge.
- Bauer, L. (1994). *Watching English Change*. Longman, London.
- Biber, D., Johansson, S., Conrad, S., and Finnegan, E. (1999). *Longman Grammar of Spoken and Written English*. Pearson Education, Ltd., Harlow, England.
- Bos, J. (2009). Computing Genitive Superlatives. In *Proceedings of Eighth International Workshop on Computational Semantics IWCS-8*, Tilburg University, Netherlands.
- Bos, J., Clark, S., Steedman, M., Curran, J. R., and Hockenmaier, J. (2004). Wide-coverage semantic representations from a CCG parser. In *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, page 1240, Morristown, NJ, USA. Association for Computational Linguistics.
- Bos, J. and Nissim, M. (2006). An Empirical Approach to the Interpretation of Superlatives. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 9–17, Sydney, Australia. Association for Computational Linguistics.
- Carenini, G., Ng, R. T., and Zwart, E. (2005). Extracting knowledge from evaluative text. In *K-CAP '05: Proceedings of the 3rd international conference on Knowledge capture*, pages 11–18, New York, NY, USA. ACM.
- Carroll, J., Minnen, G., and Briscoe, T. (1999). Corpus annotation for parser

- evaluation. In *In Proceedings of the EACL workshop on Linguistically Interpreted Corpora (LINC)*, pages 35–41.
- Clark, S. and Curran, J. R. (2004a). The importance of supertagging for wide-coverage CCG parsing. In *Proceedings of COLING-04*, page 282288.
- Clark, S. and Curran, J. R. (2004b). Parsing the WSJ using CCG and log-linear models. In *ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 103, Morristown, NJ, USA. Association for Computational Linguistics.
- Corver, N. and Matushansky, O. (2006). At our best when at our boldest. Handout, TIN-dag.
- Curran, J. R. and Clark, S. (2003a). Investigating GIS and smoothing for maximum entropy taggers. In *Proceedings of the 10th Meeting of the EACL*, page 9198.
- Curran, J. R. and Clark, S. (2003b). Language independent NER using a maximum entropy tagger. In *Proceedings of CoNLL-03*, page 164167.
- Denoyer, L. and Gallinari, P. (2006). The Wikipedia XML Corpus. *SIGIR Forum*.
- Ganapathibhotla, M. and Liu, B. (2008). Mining Opinions in Comparative Sentences. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 241–248. Coling 2008 Organizing Committee.
- Grice, H. P. (1975). Logic and Conversation. In *Syntax and Semantics*.
- Grover, C. and Tobin, R. (2006). Rule-Based Chunking and Reusability. In *Proceedings of LREC 2006 (Conference on Language Resources and Evaluation)*.
- Heim, I. (1985). Notes on comparatives and related matters.
- Heim, I. (1999). Notes on Superlatives.
- Hildebrandt, W., Katz, B., and Lin, J. (2004). Answering definition questions using multiple knowledge sources. In *In Proceedings of HLT-NAACL 2004*, pages 49–56. HLT.
- Hu, M. and Liu, B. (2004). Mining Opinion Features in Customer Reviews. In *Proceedings of AAAI 2004*.
- Huddleston, R. and Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press, New York.

- Janoschka, A. (2004). *Web Advertising: New Forms of Communication on the Internet*. John Benjamins Publishing Company, Amsterdam.
- Jijkoun, V. and de Rijke, M. (2006). Overview of the WiQA task at CLEF 2006. In *CLEF*, pages 265–274.
- Jindal, N. and Liu, B. (2006a). Identifying comparative sentences in text documents. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 244–251, New York, NY, USA. ACM.
- Jindal, N. and Liu, B. (2006b). Mining Comparative Sentences and Relations. In *Proceedings of AAAI 2006*.
- Kaisser, M., Scheible, S., and Webber, B. (2006). Experiments at the University of Edinburgh for the TREC 2006 QA track. In *Proceedings of TREC 2006*.
- Kennedy, C. (1997). Comparison and Polar Opposition. In *Proceedings of SALT VII*, pages 240–257. CLC Publications.
- Kim, S.-M. and Hovy, E. (2004). Determining the sentiment of opinions. In *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, page 1367, Morristown, NJ, USA. Association for Computational Linguistics.
- Klein, E. (1980). A Semantics for Positive and Comparative Adjectives. *Linguistics and Philosophy*, 4(1):1–45.
- Knott, Alistair, M. O. J. O. and Mellish, C. (1997). Defeasible Rules in Content Selection and Text Structuring. In *In Proceedings of the 6th European Workshop on Natural Language Generation, Gerhard-Mercator University*, pages 2–6.
- Lehrer, A. (1974). *Semantic Fields and Lexical Structure*. North-Holland, Amsterdam.
- Lin, J. and Demner-Fushman, D. (2006). Will pyramids built of nuggets topple over? In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 383–390, Morristown, NJ, USA. Association for Computational Linguistics.
- Linden, K. (2008). What's the Most...? Answering Superlative Questions. Master's

- thesis, School of Philosophy, Psychology and Language Sciences, University of Edinburgh.
- Lyons, J. (1977). *Semantics*. Cambridge University Press, Cambridge.
- Maedche, A. and Staab, S. (2000). Mining Ontologies from Text. In *EKAW '00: Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management*, pages 189–202, London, UK. Springer-Verlag.
- McKeown, K. R. (1985). *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press, Cambridge.
- Merlo, P. and Ferrer, E. E. (2006). The notion of argument in PP attachment. *Computational Linguistics*, 32(2):341–378.
- Merlo, P. and Leybold, M. (2001). Automatic distinction of arguments and modifiers: the case of prepositional phrases. In *In Proceedings of the Fifth Computational Natural Language Learning Workshop (CoNLL2001)*, pages 121–128.
- Milosavljevic, M. (1999). *The Automatic Generation of Comparisons in Descriptions of Entities*. PhD thesis, Microsoft Research Institute, Macquarie University, Sydney, Australia.
- Minnen, G., Carroll, J., and Pearce, D. (2001). Applied morphological processing of English. *Natural Language Engineering*, 7(3):207–223.
- Navigli, R., Velardi, P., and Gangemi, A. (2003). Ontology Learning and its Application to Automated Terminology Translation. *IEEE Intelligent Systems*, 18(1):22–31.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.
- Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pages 79–86, Morristown, NJ, USA. Association for Computational Linguistics.
- Popescu, A.-M. and Etzioni, O. (2005). Extracting product features and opinions from reviews. In *HLT '05: Proceedings of the conference on Human Language*

- Technology and Empirical Methods in Natural Language Processing*, pages 339–346, Morristown, NJ, USA. Association for Computational Linguistics.
- Prager, J. M., Chu-Carroll, J., and Czuba, K. (2001). Use of WordNet Hypernyms for Answering What-Is Questions. In *TREC*.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.
- Razmara, M. and Kosseim, L. (2007). A little known fact is... Answering Other questions using interest-markers. In *Proceedings the 8th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING-2007)*, Lecture Notes in Computer Science, Volume 4394, pages 518–529, Mexico City, Mexico. Springer-Verlag.
- Reiter, E. (1990). The Computational Complexity of Avoiding Conversational Implicatures. In *ACL*, pages 97–104.
- Reiter, E. and Dale, R. (1992). A Fast Algorithm for the Generation of Referring Expressions. In *COLING*, pages 232–238.
- Ross, J. R. (1964). A partial grammar of English superlatives. Master's thesis, University of Pennsylvania. unpublished.
- Scheible, S. (2006). A Computational Treatment of Superlatives. Ph.D. Proposal, University of Edinburgh.
- Scheible, S. (2007). Towards a Computational Treatment of Superlatives. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, Prague, Czech Republic.
- Schutz, A. and Buitelaar, P. (2005). RelExt: A Tool for Relation Extraction in Ontology Extension. In *4th International Semantic Web Conference*, pages 593–606.
- Szabolsci, A. (1986). Comparative Superlatives. In Fukui, N., Rapoport, T. R., and Sagey, E., editors, *Papers in Theoretical Linguistics*, pages 245–265. MIT Working Papers, Cambridge, MA.
- van Deemter, K. (2006). Generating Referring Expressions that involve Gradable Properties. *Computational Linguistics*.

Wilson, T., Wiebe, J., and Hwa, R. (2004). Just How Mad Are You? Finding Strong and Weak Opinion Clauses. In *Proceedings of AAAI*, pages 761–769.

Yu, H. and Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136, Morristown, NJ, USA. Association for Computational Linguistics.